SECOND EDITION

Science and Pseudoscience in Clinical Psychology

edited by
Scott O. Lilienfeld
Steven Jay Lynn
Jeffrey M. Lohr

ebook

THE GUILFORD PRESS

SCIENCE AND PSEUDOSCIENCE IN CLINICAL PSYCHOLOGY

Also Available

Psychological Science in the Courtroom:
Consensus and Controversy
Edited by Jennifer L. Skeem,
Kevin S. Douglas, and Scott O. Lilienfeld

Theories of Hypnosis:
Current Models and Perspectives
Edited by Steven Jay Lynn
and Judith W. Rhue

Science and Pseudoscience in Clinical Psychology

SECOND EDITION

Edited by
Scott O. Lilienfeld
Steven Jay Lynn
Jeffrey M. Lohr

Foreword by Carol Tavris

THE GUILFORD PRESS
New York London

A Division of Guilford Publications, Inc.
72 Spring Street, New York, NY 10012
www.guilford.com

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Science and pseudoscience in clinical psychology / edited by Scott O. Lilienfeld, Steven Jay Lynn, Jeffrey M. Lohr ; foreword by Carol Tavris. — Second edition.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-4625-1789-3 (pbk. : alk. paper) — ISBN 978-1-4625-1751-0 (hardcover : alk. paper)
I. Lilienfeld, Scott O., 1960–, editor. II. Lynn, Steven J., editor. III. Lohr, Jeffrey M., editor.
[DNLM: 1. Psychology, Clinical—methods. 2. Psychotherapy—methods. 3. Research. WM 105]
RC467
616.89—dc23

2014036095

About the Editors

Scott O. Lilienfeld, PhD, is Professor of Psychology at Emory University. He is Associate Editor of the *Journal of Abnormal Psychology* and serves on the editorial boards of several other journals. Dr. Lilienfeld's principal interests are the causes and assessment of personality disorders, especially psychopathy; dissociative disorders; psychiatric classification and diagnosis; evidence-based practice in psychology; philosophical psychology; and the application of scientific thinking to psychology education. He is a recipient of the James McKeen Cattell Award for Distinguished Achievements in Applied Psychological Science from the Association for Psychological Science and serves as President of the Society for the Scientific Study of Psychopathy.

Steven Jay Lynn, PhD, ABPP, is Distinguished Professor of Psychology at Binghamton University, The State University of New York (SUNY), where he is Director of the Psychological Clinic. He has published more than 300 articles, books, and book chapters on topics including psychotherapy, hypnosis, science versus pseudoscience, psychopathology, and memory, and his research is widely cited in the media. Dr. Lynn is Founding Editor and Editor of the American Psychological Association journal *Psychology of Consciousness*. He is a recipient of the SUNY Chancellor's Award for Excellence in Scholarship and Creative Activities.

Jeffrey M. Lohr, PhD, is Professor of Psychology at the University of Arkansas–Fayetteville, where he has been on the faculty since 1975. His research interests include affective processes in anxiety and related disorders and pseudoscience in applied and clinical psychology. In the latter domain, he focuses primarily on the empirical analysis of treatment efficacy and the promotion of "fringe" treatments, especially as they relate to trauma and anxiety disorders. Dr. Lohr is Associate Editor of *The Behavior Therapist*.

Contributors

Manuel Barrera Jr., PhD, Psychology Department, Arizona State University, Tempe, Arizona

Patricia A. Boyle, PhD, Rush Alzheimer's Disease Center, Armour Academic Center, Chicago, Illinois

Kristy L. Dalrymple, PhD, Rhode Island Hospital, Providence, Rhode Island

Brett Deacon, PhD, School of Psychology, University of Wollongong, Wollongong, New South Wales, Australia

Grant J. Devilly, PhD, School of Applied Psychology and Griffith Health Institute, Griffith University (Mt Gravatt Campus), Mount Gravatt, Queensland, Australia

Trevor A. Fronius, MA, WestEd, Woburn, Massachusetts

Howard N. Garb, PhD, Psychology Research Service, Lackland Air Force Base, San Antonio, Texas

Brandon A. Gaudiano, PhD, Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University, and Psychosocial Research Program, Butler Hospital, Providence, Rhode Island

Jennifer M. Gillis, PhD, BCBA-D, Institute for Child Development, Binghamton University, The State University of New York, Binghamton, New York

Richard Gist, PhD, Kansas City (Missouri) Fire Department, Kansas City, Missouri

Russell E. Glasgow, PhD, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, Maryland

Joshua C. Gray, BA, Department of Psychology, University of Georgia, Athens, Georgia

Sarah Guckenburg, MPH, WestEd, Woburn, Massachusetts

Tammy R. Hammond Natof, PhD, private practice, Lexington, Kentucky

Meghan E. Hollis-Peel, PhD, School of Criminal Justice, Michigan State University, East Lansing, Michigan

John Hunsley, PhD, CPsych, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada

Irving Kirsch, PhD, Department of Psychology, University of Hull, Hull, United Kingdom

Elisa Krackow, PhD, Department of Psychology, West Virginia University, Morgantown, West Virginia

Catherine M. Lee, PhD, Faculty of Social Sciences, University of Ottawa, Ottawa, Ontario, Canada

Scott O. Lilienfeld, PhD, Department of Psychology, Emory University, Atlanta, Georgia

Timothy G. Locke, private practice, Brookfield, Connecticut

Elizabeth F. Loftus, PhD, School of Law, University of California, Irvine, Irvine, California

Jeffrey M. Lohr, PhD, Department of Psychological Science, University of Arkansas–Fayetteville, Fayetteville, Arkansas

Steven Jay Lynn, PhD, ABPP, Psychological Clinic, Psychology Department, Binghamton University, The State University of New York, Binghamton, New York

Pamela MacDougall, BA, WestEd, Woburn, Massachusetts

James MacKillop, PhD, Experimental and Clinical Psychopharmacology Laboratory, Department of Psychology, and Institute for Behavioral Research, University of Georgia, Athens, Georgia

Joseph T. McCann, PsyD, United Health Services Hospitals, Binghamton, New York

Jean Mercer, PhD, Department of Psychology, Richard Stockton College, Moorestown New Jersey

Timothy E. Moore, PhD, C Psych, Department of Psychology, Glendon College, York University, Toronto, Ontario, Canada

Anthony Petrosino, PhD, Learning Innovations at WestEd, Woburn, Massachusetts

Monica Pignotti, PhD, College of Social Work, Florida State University, Tallahassee, Florida

Raymond G. Romancyzk, PhD, Psychology Department, Binghamton University, The State University of New York, Binghamton, New York

Gerald M. Rosen, PhD, Department of Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Seattle, Washington

Melina Sevlever, MS, Department of Psychology, Auburn University, Auburn, Alabama

Kelley L. Shindler, MS, Psychology Department, Binghamton University, The State University of New York, Binghamton, New York

Whitney Taylor, BA, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada

Bruce A. Thyer, PhD, LCSW, BCBA-D, College of Social Work, Florida State University, Tallahassee, Florida

Laura B. Turner, MS, BCBA, Psychology Department, Binghamton University, The State University of New York, Binghamton, New York

Tracey Varker, PhD, Centre for Posttraumatic Mental Health, University of Melbourne, Melbourne, Victoria, Australia

Harald Walach, PhD, Samueli Institute, European Office, European University, Frankfurt, Germany

Daniel A. Waschbusch, PhD, Division of Child and Adolescent Psychiatry, Penn State Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, Pennsylvania

James G. Waxmonsky, MD, Division of Child and Adolescent Psychiatry, Penn State Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, Pennsylvania

Lauren M. Weinstock, PhD, Psychosocial Research Program, Butler Hospital, Providence, Rhode Island

James M. Wood, PhD, Department of Psychology, University of Texas at El Paso, El Paso, Texas

FOREWORD

The Scientist–Practitioner Gap

Revisiting "A View from the Bridge" a Decade Later

I was sitting in a courtroom, watching the title of this book—*Science and Pseudoscience in Clinical Psychology*—in action. A pediatric psychologist, a woman with a PhD in clinical psychology from a prestigious university, was testifying about the reasons for her sure and certain diagnosis that the accused woman was a "Munchausen by proxy" mother, and that the woman's teenage son was not in fact ill with an immune disorder but rather was "in collusion" with his disturbed mother to produce his symptoms. Let's call this expert Dr. M.

No one disputes that some mothers have induced physical symptoms in their children and subjected them to repeated hospitalizations; some cases have been captured on hospital video cameras. There is a term for this cruel behavior; we call it child abuse. When the child dies at the hands of an abusive parent, we have a term for that, too; we call it murder. But many clinicians suffer from syndromophilia. They have never met a behavior they can't label as a mental disorder. One case is an oddity, two is a coincidence, and three is an epidemic.

Once a syndrome is labeled, it spawns experts who are ready and willing to identify it, treat it, and train others to be ever alert for signs of it. No new disorder is "rare" to these experts; it is "mistaken" for something else or "underdiagnosed." When the new diagnosis is received with skepticism and controversy, a common reaction is not to reject it but to rename it: Munchausen by proxy (MBP) became "factitious disorder by proxy" in the DSM-IV's appendix—the interim limbo for contentious labels—and was promoted to "factitious disorder imposed on another" in DSM-5 (American Psychiatric Association, 2013). When multiple personality disorder

(MPD) officially entered DSM-III in 1980, setting off a hysterical epidemic of cases that grew into the thousands throughout the 1980s and 1990s, it wasn't science or skepticism that ended this psychiatric folly, but malpractice suits. DSM-IV took no stand, but quietly renamed the diagnosis "dissociative identity disorder." It remains in DSM-5 (see Lilienfeld & Lynn, Chapter 5, this volume).

When I was writing the foreword to the first edition in 2003, it seemed that Munchausen by proxy, or by any other name, would become the latest trendy disorder to capture clinical and media attention (Mart, 1999). Mercifully, it didn't, but Dr. M's testimony in the courtroom that day revealed the pseudoscientific assumptions, methods, and ways of thinking that are still too common in clinical practice, as other chapters in this volume will consider in depth:

- Dr. M relied on projective tests to determine whether or not the mother had psychological problems, apparently unaware that these tests have serious problems of reliability and validity (see Hunsley, Lee, Wood, & Taylor, Chapter 3, this volume). Moreover, what could those tests reveal? Evidence of a "mental disorder" in this defendant would not reliably indicate that she was an MBP mother. For that matter, all too many abusive parents have no discernible "mental disorder."

- Dr. M knew nothing about the importance of testing clinical assumptions empirically, let alone of operationally defining her terms. What does "in collusion" mean? How does an MBP mother's behavior differ from that of any mother of a chronically sick child, or, for that matter, from that of any loving mother?

- Dr. M knew nothing about the confirmation bias or the principle of falsifiability, and how these might affect clinical diagnosis (Tavris & Aronson, 2007). Once she decided this mother was a Munchausen case, that was that. Nothing the mother did or said could change her mind. This is because, she testified, Munchausen mothers are so deceptively charming, educated, and persuasive. Nothing the child said could change her mind. This is because, she said, the child naturally wants to remain with his mother, in spite of her abusiveness. No testimony from immunologists that the child really did have an immune disorder could change her mind. This is because, she explained, Munchausen mothers force doctors to impose treatments on their children by interpreting "borderline" medical conditions as problems needing intervention.

- Dr. M understood nothing about the social psychology of diagnosis—for example, how a rare problem, such as "dissociative identity disorder" or "Munchausen by proxy" syndrome, becomes overreported when clinicians start looking for it everywhere and are rewarded with fame, acclaim, and income when they find it (Acocella, 1999).

• Dr. M understood nothing about the problem of error rates (Mart, 1999): that in their zeal to avoid false negatives (failing to identify mothers who are harming their children), clinicians might significantly boost the rate of false positives (mistakenly labeling mothers as having MBP syndrome). "This disorder destroys families," she said, without apparently pausing to consider that mistaken diagnoses do the same.

In short, Dr. M managed to get a PhD in clinical psychology without having acquired a core understanding of the basic principles of scientific thinking.

Is Dr. M an anachronism, doomed to extinction by the rise of evidence-based practice in medicine and empirically supported therapies in psychology? I suspect that many clinical scientists would say yes. After all, the establishment of the Psychological Clinical Science Accreditation System (PCSAS) is surely evidence of a sea change in clinical training: a nationwide effort to improve low training standards in graduate programs by identifying and accrediting outstanding clinical graduate programs that train high-quality researchers. PCSAS is certainly an important step forward in closing the gap. But, as Lilienfeld (2013) has observed, "PCSAS recognizes only research-oriented programs; it neglects practitioner-oriented programs that are doing a solid job of training their students to think and practice scientifically. As the PCSAS initiative progresses, we must remain cognizant of the pressing need to train clinicians to think and operate scientifically."

Dr. M would still be at home in most graduate clinical psychology programs and psychiatric residencies today, where students can earn a PhD or an MD without ever having considered the basic epistemological assumptions and methods of their profession: What kinds of evidence are needed before we can draw strong conclusions? Are there alternative hypotheses that I have not considered? Why are so many diagnoses of mental illness based on consensus—a group vote—rather than on empirical evidence, and what does this process reveal about problems of reliability and validity in diagnosis? An ethnographic study of the training of psychiatrists showed that psychiatric residents learn how to make quick diagnoses, prescribe medication, and, in a dwindling number of locations, do psychodynamic talk therapy, but rarely do they learn to be skeptical, ask questions, analyze research, or consider alternative explanations or treatments (Luhrmann, 2000). Psychiatric training has not improved in the decade since; on the contrary, the move toward psychiatrist-as-medication-dispenser has escalated (Kirk, Gomory, & Cohen, 2013).

Moreover, despite decades of controlled research that have shown which therapies are most efficacious for specific problems, most of the results have not yet filtered down to the great majority of practitioners. Large numbers of people with anxiety disorders do not receive exposure-based methods, and most children with autism spectrum disorders are

receiving unsupported interventions (Lilienfeld, 2013; Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013). Most people with depression routinely are given antidepressants but not cognitive-behavioral therapy, and, as meta-analyses repeatedly find, antidepressants are largely placebic for people with all but the severest form of depression (Kirsch et al., 2008; Fournier et al., 2010; see Walach & Kirsch, Chapter 12, this volume). Most dishearteningly, large numbers of clinicians continue to use suggestive techniques, such as hypnosis and guided imagery, to recover allegedly repressed memories of abuse (Baker, McFall, & Shoham, 2008; Lilienfeld et al., 2013).

Scott Lilienfeld and colleagues (2013) have identified several sources of clinicians' resistance to evidence-based practice. One is that many therapists want to believe that if the client changes, it was because of what the therapist did, thereby overlooking other factors in the client's life that might have contributed to improvement. Many therapists still hold the psychodynamic belief that early childhood experience is *the* cause of adult emotional problems. Many don't understand complex statistics or know how to interpret the literature on outcomes of psychotherapy (and hence don't read the journals). Many still have the view that it is up to skeptics to prove their position that some therapy *doesn't* help, rather than up to proponents of that therapy to show that it *does*. Most or all of these sources of resistance may be subsumed under the banner of self-justification: The way I do things is the right way, and if you tell me it's the wrong way and I need to learn some statistics, you are insulting my competence and professionalism and deep understanding of human nature, so the hell with you (Tavris & Aronson, 2007).

For their part, the general public is no better informed about advances in clinical science. Because psychotherapists tend to be the ones who are writing advice columns, writing pop-psych books, going on talk shows, and testifying as experts in court cases, the public is largely ignorant of the kind of research done by psychological scientists on clinical issues. Here are just a few of the widely held beliefs, promoted by many psychotherapists, that have been discredited by empirical evidence and that are dispelled in the pages of this edited volume:

- Almost all abused children become abusive parents.
- Almost all children of alcoholics become alcoholic.
- Children never lie about sexual abuse.
- Childhood trauma invariably produces emotional symptoms that carry on into adulthood.
- Memory works like a tape recorder, clicking on at the moment of birth.
- Hypnosis can reliably uncover buried memories.
- Traumatic experiences are usually repressed.
- Hypnosis reliably uncovers accurate memories.

- Subliminal messages influence behavior.
- Children who masturbate or "play doctor" have probably been sexually molested.
- If left unexpressed, anger builds up like steam in a teapot until it explodes in verbal or physical aggression.
- Projective tests like the Rorschach validly diagnose personality disorders, most forms of psychopathology, and sexual abuse.

All of these widely held but mistaken ideas can have, and have had, devastating consequences in people's lives. I once heard a social worker explain to a judge why she had decided to remove a child from her mother's custody: The mother had been abused as a child, and "we all know," she said, that that is a major risk factor for the mother's abuse of her own child one day. Obviously, no one had taught this social worker about disconfirming cases. In fact, researchers have known for decades that while being abused as a child increases the risk of becoming an abusive parent, the large majority of abused children—about two-thirds—do not become abusive parents (Kaufman & Zigler, 1987).

Of course, the gap between psychological science and clinical practice has always existed. In many ways, it is no different from the natural tensions that exist between researchers and practitioners in any field—medicine, engineering, education, psychiatry, physics—when one side is doing research and the other is working in an applied domain: Their goals and training are inherently different. Medical practice faces two kinds of error: Many physicians continue to rely on outmoded methods and beliefs and won't change the techniques they learned in medical school; many others are too quick to replace tried-and-true methods or medicines with newer, as-yet-unvalidated ones that come along every week with fanfare, direct-to-consumer advertising, and attractive pharmaceutical reps.

The goal of psychotherapy (as of medicine) is to help the suffering individual who is sitting there; the goal of science is to explain and predict the behavior of people in general. That is another reason that many psychotherapists have protested the very effort to measure therapy outcomes. Research methods and findings, they maintain, capture only a small, shriveled image of the real person (e.g., Edelson, 1994). You cannot quantify the complexity of the complaints that people bring to treatment, let alone those that emerge *during* treatment. You often cannot quantify what people want from treatment either, apart from an overall "Please reduce my ache," partly because clients don't always know what they want: Improvement of symptoms? Justification of behavior? A supportive listener? Validation? Healing a bereaved heart? Getting Harry or Harriet to *listen* to me, for chrissakes? Therapy, many clinicians argue, was helping people long before science or psychology was invented. Professional training, therefore, should teach students how to do therapy wisely and well, not how to do science.

In psychology, this divergence in goals and training was present at its conception. Empirical psychology and psychoanalysis were born of different fathers in the late 19th century and never got along. Throughout the 20th century, they quarreled endlessly over fundamental assumptions about the meaning of science and truth. How do we know what is true? What kind of evidence is required to support a hypothesis? To early psychoanalysts, "science" had nothing to do with controlled experiments, interviews, or statistics (Hornstein, 1992). In constructing what they saw as a "science of the mind," psychoanalysts relied solely on their own interpretations of cases they saw in therapy, of myths and literature, and of people's behavior. To empirically minded psychologists, the idea that analysts could claim to be doing science while chucking out the cardinal rules of the scientific method—replicable findings, verifiable data, objective confirmation of evidence, and the concerted effort to control prejudices and any other possible sources of bias—was alarming. When psychoanalysis first became popular in the United States in the 1920s, many scientific psychologists regarded it as a popular craze, something on a par with mind reading or phrenology, which would blow over. John Watson called it "voodooism." "Psychoanalysis attempts to creep in wearing the uniform of science," wrote another critic at the time, "and to strangle it from the inside" (quoted in Hornstein, 1992, p. 476). Replace *psychoanalysis* in that sentence with rebirthing or thought field therapy (TFT), and the attitude is just as prevalent today among psychological scientists.

By the 1960s and 1970s, as the popularity of psychoanalysis was waning, newer therapies were emerging. It was easy to tell how pseudoscientific *they* were. Unlike the Freudians, who said you needed to be in treatment for 5 years, these new guys were offering miracle therapies that promised to cure you in 5 days, 5 minutes, or 5 orgasms.

In the heyday of the countercultural revolution, these therapies multiplied like rabbits. Martin Gross's (1978) book *The Psychological Society* included marathon therapy, encounter therapy, nude therapy, crisis therapy, primal-scream therapy, electric sleep therapy, body-image therapy, deprivation therapy, expectation therapy, alpha-wave therapy, "art of living" therapy, "art of loving" therapy, and "do it now" therapy. By the 1980s, pop therapy had gone high-tech. Electrical gizmos promised to get both halves of your brain working at their peak (Chance, 1989): the Graham Potentializer, the Tranquilite, the Floatarium, the Transcutaneous Electro-Neural Stimulator, the Brain SuperCharger, and the Whole Brain Wave Form Synchro-Energizer.

At first, most psychological scientists paid as little heed to the explosion of post-Freudian pop therapies and technologies as they had to psychoanalysis. These therapies were a blot on the landscape of psychology, perhaps, but a benign nuisance; the worst thing that consumers might suffer was a loss of money and dignity.

But by the mid-1980s North America was in the midst of three social contagions, which some call hysterical epidemics or moral panics (Jenkins, 1998; Showalter, 1997): recovered memory therapy, the daycare sex-abuse scandals, and multiple personality disorder. All three had been fomented by the erroneous and scientifically unvalidated claims of psychotherapists, using subjective and unreliable methods. Moreover, many of the people making these claims were psychiatrists and clinical psychologists, along with social workers and therapists of various stripes who had taken a weekend course somewhere on child abuse. Hadn't they taken Psychology 101? Had no one taught them about control groups, memory, child development, the limitations of hypnosis? Apparently not. Poole, Lindsay, Memon, and Bull (1995) found that large minorities of registered psychotherapists in the United States and England were using subjective, highly influential techniques such as hypnosis, dream analysis, and guided imagery related to abuse situations to "uncover" repressed memories of childhood sexual abuse. As noted previously, the situation has not improved much in many quarters (Lilienfeld et al., 2013).

The recovered memory movement revealed in glaring lights how far apart empirical psychology and clinical psychology had grown. After World War II, the two sides had tried forging an alliance: The "scientist–practitioner model" would govern the training of clinical psychologists, who would draw on the most relevant findings of research psychology in diagnosing and treating clients. This harmonious ideal did not last long; like all ideals, it was easier to praise than to follow. The inherent tensions between the two sides grew, and by the early 1990s, researchers and clinicians were speaking openly of the "scientist–practitioner gap" (Persons, 1991).

Calling it a "gap," however, is like saying there is an Israeli–Arab "gap" in the Middle East. It is a simmering conflict that occasionally erupts into outright warfare, involving deeply held beliefs, political passions, views of human nature and the nature of knowledge, and—as all wars ultimately involve—money, territory, and livelihoods. Anyone who has disputed or publicly questioned the accuracy of recovered memories of sex abuse, the concepts of repression and dissociation, or any of the many popular but unvalidated therapies (e.g., facilitated communication [FC], critical incident stress debriefing [CISD], or rebirthing and other attachment therapies) knows the inflammatory nature of such criticism and the invective with which it will often be received.

The current war between psychological scientists and clinicians—as opposed to the normal grumbling and intellectual debate that had been going on for decades—stems from several economic and cultural forces. One has been the rapid proliferation of psychotherapists of all kinds. Many are graduated from "freestanding" schools, unconnected to university psychology departments, where they typically learn only to do

therapy—and sometimes only a vague kind of psychodynamic therapy at that. Others take brief certification courses in hypnotherapy or various counseling programs, and then promote themselves as experts in a particular method. Because so many kinds of therapy are now competing in the marketplace of treatments, and because of the economic challenges posed by managed care, these specialties have become precious sources of income to many therapists. People who earn their livings from giving Rorschach workshops, TFT training, setting up crisis intervention programs, administering projective tests, or diagnosing sexual abuse are not going to be receptive to evidence questioning the validity of their methods or assumptions.

In North America today, entire industries sail under the flags of pseudoscience, and there is a cultural reason for their popularity as well as an economic one. Cross-cultural psychologists have studied how cultures differ in their need for certainty and tolerance of ambiguity, and hence, for example, whether they are willing to try to live with life's inherent uncertainties or pass laws to try to reduce or eliminate them (Cvetkovich & Earle, 1994; Hofstede & Bond, 1988). The United States is a culture that has a low tolerance for uncertainty; hence our attraction to zero-tolerance policies that fruitlessly attempt to eradicate drug abuse and to abstinence-only sex education programs that attempt to eradicate sex among teenagers. Oh, sure; go for it.

In such a culture, pseudoscience is particularly attractive because pseudoscience by definition promises certainty, whereas science gives us probability and doubt. Pseudoscience is popular because it confirms what we believe; science is unpopular because it makes us question what we believe. Good science, like good art, often upsets our established ways of seeing the world. Bruce Rind and his colleagues discovered this to their dismay when they published their meta-analysis suggesting that child sexual abuse, carefully defined, does not inevitably produce severe psychopathology in adulthood (Rind, Tromovitch, & Bauserman, 1998). Did the public rise as one to praise them for this scientific reassurance that most people survive terrible experiences? Hardly. Instead, Congress passed a resolution condemning their research, and an odd consortium of religious conservatives and recovered memory psychotherapists mobilized an attack on the researchers' motives, their methods, and their findings (Rind, Tromovitch, & Bauserman, 2000).

Longing for certainty about difficult problems, the public turns to psychologists who will give them *the* answer: Which parent should get custody? Is this rapist cured? Is this child's terrible accusation accurate? How can we prevent murderous rampages by school shooters? What therapy can make me better, fast? Scientists speak in the exasperating language of probability: "It is likely that . . . ." How much more appealing are the answers of clinicians who are prepared to say, with certainty, "This mother is paranoid, believing that her husband is out to get her"; "This rapist is definitely

cured"; "Children never lie about sexual abuse"; "Thought field therapy can fix you in 5 minutes."

Pseudoscientific programs, potions, and therapies have always been an entrenched part of American culture, along with moonshine and Puritanism. The cultural mix of pragmatism, an optimistic belief that anything can be changed and improved, and impatience with anything that takes much time has created a longstanding market for instant solutions. All a clever entrepreneur has to do is apply a formula historically guaranteed to be successful: (Quick Fix + Pseudoscientific Gloss) × Credulous Public = High Income. *The Tapping Solution* is a 2013 exemplar: "Tapping, also known as EFT, is a powerful tool for improving your life on multiple levels," boasts the Amazon description: "mental, emotional, and physical. It has been proven to effectively address a range of issues—from anxiety, chronic pain, addiction, and fear to weight control, financial abundance, stress relief, and so much more. It's also one of the easiest and fastest practices to learn. You can learn it in minutes, do it anywhere and on virtually any issue, and oftentimes experience immediate results" (see Pignotti & Thyer, Chapter 7, this volume).

Well, wow! Why bother with pesky data, let alone with actually talking to another human being? Yet we can be sure that when EFT, TFT, FC, NLP, and their acronymic siblings have traveled the route of electric sleep therapy and the Transcutaneous Electro-Neural Stimulator, new miracle therapies with different acronyms will rise to take their place. It's the American way.

Pseudoscientific therapies will always remain with us because so many economic and cultural interests are promoting them. But their potential for harm to individuals and society is growing, which is why it is more important than ever for psychological scientists to expose their pretensions and dangers. As Richard McNally is fond of saying, the best way to combat pseudoscience is to do good science. Indeed, good psychological science has already helped slow, if not yet overturn, the hysterical epidemics of our recent history that wrought so much harm. Psychological science has given us a better understanding of memory, of the processes of influence and suggestibility in therapy that create such iatrogenic disorders as "multiple personalities," and of better ways of interviewing children and assessing their accounts and memories. Good psychological science has helped clinicians develop effective interventions for a wide array of specific problems, from specific phobias to binge eating. Research has distinguished therapeutic techniques that are merely ineffective from those that are harmful, such as "rebirthing," in which therapists in Colorado smothered a 10-year-old girl to death as they supposedly helped her to be "reborn," and CISD programs, which can actually delay a victim's recovery from disasters and traumas (see Lohr, Gist, Deacon, Devilly, & Varker, Chapter 10, this volume).

Yet the essential difference between scientific psychology and psychotherapy will always remain, too. "In therapy, the trick is to tell stories that

satisfy; in science the trick is to tell stories that predict," says Michael Nash (personal communication). "A story that is satisfying—a compelling narrative that makes our lives meaningful—need not be true in some objective sense. So therapists are right when they say that research can't help individuals learn to live with suffering, resolve moral dilemmas, or make sense of their lives. But they must be disabused of the notion that their clients' stories are literally true, or that they have no part in shaping them."

Exactly. To this difficult challenge I would add another, one that in the past decade has become increasingly important for all clinicians to understand: In this age of the biomedical–industrial complex, the pressure is enormous to equate "science" with biology and medicine, thereby reducing human laments, woes, and problems to faulty brain circuits or hormone levels. Today it is not enough to worry about quick-fix therapies; we need also to worry about quick-fix medication. In the aftermath of publication of DSM-5, which was greeted with a chorus of criticism (Frances, 2013; Greenberg, 2013; Kirk, Gomory, & Cohen, 2013), the National Institute of Mental Health (NIMH) announced that it was developing a different way of classifying mental problems that would be less vulnerable to subjective opinion. Great goal, wrong turn. The NIMH effort is based on the assumption that mental disorders reflect abnormalities in brain circuits and other biological phenomena, and it plans to produce "precision medicine" to treat them. What "precision medicine" will help Harriet hear Harry's heartache? Why should anyone take a "precision medicine" for a phobia that can be effectively treated with exposure? And what "precision medicine" will be prescribed for the side effects of the first one?

In sum, we need the research and scientific thinking of clinical scientists *and* the intelligent practice of psychotherapy more than ever. The science–practice gap may not matter much in the subjective, immeasurable process of helping a client find wisdom and a story that satisfies. But it does matter in the practice of incompetent, coercive, or harmful therapy. And it matters profoundly when therapists step outside their bounds, claiming expertise and *certainty* in domains in which unverified clinical opinion can ruin lives, and where knowledge of good psychological science can save them.

Carol Tavris, PhD
Los Angeles, California

REFERENCES

Acocella, J. (1999). *Creating hysteria: Women and multiple personality disorder.* San Francisco: Jossey-Bass.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Baker, T. B., McFall, R. M., & Shoham, V. (2008). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest*, *9*, 67–103.

Brainerd, C. J., Reyna, V. F., & Brandse, E. (1995). Are children's false memories more persistent than their true memories? *Psychological Science*, *6*, 359–364.

Chance, P. (1989, November). The other 90%. *Psychology Today*, pp. 20–21.

Cvetkovich, G. T., & Earle, T. C. (1994). Risk and culture. In W. J. Lonner & R. Malpass (Eds.), *Psychology and culture* (pp. 217–224). Boston: Allyn & Bacon.

Edelson, M. (1994). Can psychotherapy research answer this psychotherapist's questions? In P. F. Talley, H. H. Strupp, & S. F. Butler (Eds.), *Psychotherapy research and practice: Bridging the gap* (pp. 60–87). New York: Basic Books.

Fournier, J. C., DeRubeis, R. J., Hollon, S.D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., et al. (2010, January 6). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *Journal of the American Medical Association*, *303*, 47–53.

Frances, A. (2013). *Saving normal*. New York: William Morrow.

Garry, M., Manning, C. G., & Loftus, E. F. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin and Review*, *3*, 208–214.

Gross, M. (1978). *The psychological society*. New York: Random House.

Greenberg, G. (2013). *The book of woe*. New York: Blue Rider Press/Penguin.

Hofstede, G., & Bond, M. H. (1988). The Confucius connection: From cultural roots to economic growth. *Organizational Dynamics*, *18*, 5–21.

Hornstein, G. (1992). The return of the repressed: Psychology's problematic relations with psychoanalysis, 1909–1960. *American Psychologist*, *47*, 254–263.

Jenkins, P. (1998). *Moral panic: Changing concepts of the child molester in modern America*. New Haven, CT: Yale University Press.

Katz, Z. (2001). Canadian psychologists' education, trauma history, and the recovery of memories of childhood sexual abuse (doctoral dissertation, Simon Fraser University, 2001). *Dissertations Abstracts International*, *61*, 38–48.

Kaufman, J., & Zigler, E. (1987). Do abused children become abusive parents? *American Journal of Orthopsychiatry*, *57*, 186–192.

Kirk, S. A., Gomory, T., & Cohen, D. (2013). *Mad science: Psychiatric coercion, diagnosis, and drugs*. New Brunswick, NJ: Transaction.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine*, *5*, e45.

Lilienfeld, S. O. (2013, March). Closing the science–practice gap. *[APS] Observer*, *26*(3).

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review*, *33*(7), 883–900.

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory*. New York: St. Martin's Press.

Luhrmann, T. M. (2000). *Of two minds: The growing disorder in American psychiatry*. New York: Knopf.

Mart, E. (1999). Problems with the diagnosis of factitious disorder by proxy in forensic settings. *American Journal of Forensic Psychology*, *17*, 69–82.

Nunez, N., Poole, D. A., & Memon, A. (2003). Psychology's two cultures revisited: Implications for the integration of science with practice. *Scientific Review of Mental Health Practice*, *2*, 8–19.

Persons, J. B. (1991). Psychotherapy outcome studies do not accurately represent current models of psychotherapy: A proposed remedy. *American Psychologist*, *46*, 99–106.

Polusny, M. A., & Follette, V. M. (1996). Remembering childhood sexual abuse: A national survey of psychologists' clinical practices, beliefs, and personal experiences. *Professional Psychology: Research and Practice*, *27*, 41–52.

Poole, D. A., Lindsay, D. S., Memon, A., & Bull, R. (1995). Psychotherapy and the recovery of memories of childhood sexual abuse: U.S. and British practitioners' opinions, practices, and experiences. *Journal of Consulting and Clinical Psychology*, *63*, 426–437.

Rind, B., Tromovitch, P., & Bauserman, R. (1998). A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin*, *124*, 22–53.

Rind, B., Tromovitch, P., & Bauserman, R. (2000). Condemnation of a scientific article: A chronology and refutation of the attacks and a discussion of threats to the integrity of science. *Sexuality and Culture*, *4*, 1–62.

Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past*. New York: Basic Books.

Showalter, E. (1997). *Hystories: Hysterical epidemics and modern culture*. New York: Columbia University Press.

Tavris, C., & Aronson, E. (2007). *Mistakes were made (but not by ME): Why we justify foolish beliefs, bad decisions, and hurtful acts*. New York: Harcourt.

Preface

Reading this book may well leave readers feeling angry, or at least indignant. Judging from the feedback we have received, the first edition did. Some readers may object to portions of the book on the grounds that their preferred clinical techniques or brands of psychotherapy have been targeted for critical examination. For them, this book will be a bitter pill to swallow. Other readers may be deeply disturbed, even incensed, by the growing proliferation of questionable and unsupported techniques in clinical psychology. For them, the revised version of this book is long overdue. If we manage to leave readers from both groups at least a bit distressed, we will have been successful, because we will have gotten their attention. Just as important, we hope that this book's second edition, like its first, will generate greatly needed discussion and debate regarding a set of crucial issues in mental health practice that have often received inadequate attention.

Our purpose in this revised edition is to subject a variety of therapeutic, assessment, and diagnostic techniques in clinical psychology to incisive and impartial scientific scrutiny. We have elected to focus on techniques that are novel, controversial, or even questionable, but that are influential and widely used. By providing thoughtful evaluations of clinical techniques on the boundaries of present scientific knowledge, we intend to assist readers with the crucial goal of distinguishing science from pseudoscience in mental health practice.

As will become clear throughout the book, unscientific and otherwise questionable techniques have increasingly come to dominate the landscape of clinical psychology and allied fields. Survey data suggest that for many psychological conditions, including mood and anxiety disorders, patients are more likely to seek out and receive scientifically unsupported rather than supported interventions. Yet no book exists to help readers differentiate techniques within clinical psychology that are ineffective, undemonstrated, or harmful from those that are grounded solidly in scientific evidence.

There have been many developments since the publication of this book's first edition over a decade ago. In this revised edition, we update all of our chapters in light of extensive recent research on the efficacy of these techniques, and extend our examination of questionable science to the treatment of additional conditions, most notably attachment disorders and child and adolescent antisocial disorders (see Chapters 15 and 16). In addition, we examine how the growth of empirically supported therapies has altered the landscape of contemporary psychotherapy (see Chapter 6 in particular). Wherever relevant, chapters have been revised to accommodate recent changes in the latest edition of the DSM, which was published in May 2013.

This book is the first major volume devoted exclusively to distinguishing scientifically unsupported from scientifically supported practices in modern clinical psychology. Many readers may find this fact surprising. Nevertheless, as we point out later in the book (Chapter 17), the field of clinical psychology has traditionally been reluctant to subject novel and controversial methods to careful scientific evaluation. This reluctance has left a major gap, and to a substantial extent we hope that our book will fill it.

We have urged the authors of each chapter to be as objective and dispassionate as possible. In addition, we have encouraged them to be not only appropriately critical when necessary, but constructive. To this end, each chapter features both a discussion of which clinical techniques are ineffective, unvalidated, or undemonstrated, and also a discussion of which techniques are empirically supported or promising. Our mission is not merely to debunk—although in certain cases debunking is a needed activity in science—but to enlighten. Not all methods that are novel or superficially implausible are necessarily worthless or ineffective. Reflexive dismissal of the new and untested is as ill advised as is blind acceptance. We have tried to ensure that our authors steer clear of both errors.

This book should be of considerable interest to several audiences: (1) practicing clinicians across the spectrum of mental health professions, including clinical psychology, psychiatry, social work, counseling, family therapy, public health, and psychiatric nursing; (2) academicians and researchers whose work focuses on psychopathology and its diagnosis and treatment; (3) current and would-be consumers of mental health treatment techniques, as well as their friends and loved ones; (4) educated laypersons interested in mental illness; (5) science writers and journalists; and (6) graduate students and advanced undergraduates wishing to learn more about the science and pseudoscience of clinical psychology and allied fields. With respect to the lattermost group, this book is suitable as either a primary or a supplemental text for graduate and advanced undergraduate courses in clinical psychology, psychotherapy, and assessment. In addition, this book should be of considerable interest to attorneys, educators, physicians, nurses, policymakers, and others whose work bears on clinical psychology and allied mental health disciplines.

Although many of the chapters deal with conceptually and methodologically challenging issues, we have tried to keep technical language to a minimum. In addition, each of the major chapters of the book contains a glossary of key concepts and terms that should prove useful to readers unfamiliar with each major content area.

We are grateful to a number of individuals who have helped to bring this book to fruition. In particular, we thank Jim Nageotte, Jane Keislar, and Kitty Moore at The Guilford Press, whose advice, assistance, and moral support throughout both editions of this book have been invaluable. We also thank Richard McNally, David Tolin, James Herbert, John Ruscio, Jerry Davison, Jerry Rosen, Richard Gist, Grant Devilly, Robert Montgomery, the late John W. Bush, Liz Roemer, Ron Kleinknecht, Carol Tavris, David Faust, Sally Satel, and a host of other colleagues and friends whose ideas have helped to inform and shape this book. Their intellectual contributions, not to mention their friendship and support, have been indispensable.

Scott O. Lilienfeld
Steven Jay Lynn
Jeffrey M. Lohr

Contents

PART II. OVERARCHING CONTROVERSIES IN PSYCHOLOGICAL TREATMENT

PART III. CONTROVERSIES IN THE TREATMENT OF ADULT DISORDERS

PART IV. CONTROVERSIES IN THE TREATMENT OF CHILD AND ADOLESCENT DISORDERS

CHAPTER ONE

Science and Pseudoscience in Clinical Psychology

Initial Thoughts, Reflections, and Considerations

Scott O. Lilienfeld, Steven Jay Lynn, and Jeffrey M. Lohr

In many ways, the sprawling terrain of clinical psychology and allied disciplines (e.g., psychiatry, social work, counseling, school psychology, psychiatry nursing) houses two largely disconnected worlds. One world consists of researchers and practitioners who ground their work largely in scientific evidence. Investigators in this first world adhere to scientific methods in their research, availing themselves of these methods as crucial safeguards against biases in their inferences. Practitioners in this first world actively consume research findings and base their interventions and diagnostic methods largely on the best available published findings. The other world, which is largely unknown to many academics ensconced comfortably in their Ivory Tower, consists of mental health professionals who routinely neglect research evidence (Dawes, 1994). Many professionals in this second world are not regular consumers of scientific findings, and they commonly administer therapeutic and assessment methods that are either unsupported or inadequately tested.

Indeed, over the past several decades, clinical psychology and related disciplines have witnessed a change in the relation between science and practice. A growing minority of clinicians appear to be basing their therapeutic and assessment practices primarily on clinical experience and subjective

intuition rather than on controlled research evidence. As a consequence, the term "scientist–practitioner gap" is being invoked with increasing frequency (see the foreword to this volume by Carol Tavris; Baker, McFall, & Shoham, 2008; Fox, 1996), and concerns that the scientific foundations of clinical psychology are steadily eroding continue to be voiced in many quarters (Dawes, 1994; Kalal, 1999; McFall, 1991). Fueling these worries are surveys of clinical psychologists and other mental health professionals, which reveal that large percentages of them are skeptical of the need for evidence-based practice (Baker et al., 2008). Many report that they place considerably more weight on their clinical experience, intuition, and theoretical orientation than on controlled research evidence when selecting interventions (Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013; Pignotti & Thyer, 2009). As the history of medicine teaches us, this devaluation of scientific evidence is likely to have deleterious consequences for our clients (Grove & Meehl, 1996).

It is largely these concerns that have prompted us to compile this edited volume, which features chapters by distinguished experts across a broad spectrum of areas within clinical psychology. Given the markedly changing landscape of clinical psychology, we believe the second edition of this book to be both timely and important.

Much has changed since the publication of the first edition of this volume over a decade ago (Lilienfeld, Lynn, & Lohr, 2003). These changes make a revised edition imperative. On the one hand, there are some grounds for optimism. In the years following the appearance of the first edition, the field of clinical psychology has seen a heightened focus on evidence-based practice, accompanied by a movement to identify empirically supported therapies for specific psychological conditions. On the other hand, there are ample reasons for continuing concern. As the chapters to follow make clear, many or most domains of clinical practice continue to be plagued by the widespread use of questionable or unvalidated techniques. In this book's second edition, we have not only updated our previous chapters in light of recent research, but added two new chapters focused on clinical domains that have become a focus of increasing concern over the past decade: attachment therapies (see Mercer, Chapter 15, this volume) and questionable treatments for childhood and adolescent antisocial behaviors (see Petrosino, MacDougall, Hollis-Peel, Fronius, & Guckenberg, Chapter 16, this volume). As a consequence of these updates and additions, this volume remains the most comprehensive resource for practitioners, researchers, instructors, and students who wish to distinguish well-supported from poorly supported techniques in clinical psychology and related fields.

Some might contend that the problem of unsubstantiated treatment techniques is not new and has in fact dogged the field of clinical psychology virtually since its inception. To a substantial extent, they would be correct. Nevertheless, the growing availability of information resources (some of which have also become misinformation resources), including popular psychology books and the Internet, the apparent upsurge of mental health

training programs that do not emphasize scientific training (Baker et al., 2008; Beyerstein, 2001), and the burgeoning industry of fringe psychotherapies, have magnified the gulf between scientist and practitioner to a problem of serious, even critical, proportions.

THE SCIENTIST–PRACTITIONER GAP AND ITS SOURCES

What are the primary sources of the growing scientist–practitioner gap? As many authors have noted (see Baker et al., 2008, Gambrill, 2006, and Lilienfeld, 1998, for discussions), some practitioners in clinical psychology and related mental health disciplines continue to use unsubstantiated, untested, and otherwise questionable treatment and assessment methods. Moreover, psychotherapeutic methods of unknown or doubtful validity are proliferating on an almost weekly basis. For example, one highly selective sampling of fringe psychotherapeutic practices (Eisner, 2000; see also Singer & Lalich, 1996) included neurolinguistic programming, Thought Field Therapy, Emotional Freedom Technique, rage reduction therapy, primal scream therapy, feeling therapy, Buddha psychotherapy, past lives therapy, future lives therapy, alien abduction therapy, angel therapy, rebirthing, Sedona method, Silva method, entity depossession therapy, vegetotherapy, palm therapy, and a plethora of other methods (see also Pignotti & Thyer, Chapter 7, this volume).

Furthermore, a great deal of academic and media coverage of such fringe treatments is accompanied by scant critical evaluation. One edited volume (Shannon, 2002) features 23 chapters on largely unsubstantiated psychological techniques, including music therapy, homeopathy, breath work, therapeutic touch, aromatherapy, medical intuition, acupuncture, and body-centered psychotherapies. Nevertheless, in most chapters these techniques receive minimal scientific scrutiny (see Corsini, 2001, for a similar example). More recently, consumers can purchase volumes on energy therapies, body work therapies, and brain-based psychotherapies, among scores of others, that are largely or entirely devoid of empirical support (e.g., Feinstein, Eden, Craig, & Bowen, 2005; Heller & Duclos, 2012).

Additional threats to the scientific foundations of clinical psychology and allied fields stem from the thriving self-help industry. This industry produces hundreds of new books, manuals, and audiotapes each year (see Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume), many of which promise rapid or straightforward solutions to complex life problems. Although some of these self-help materials may be efficacious, the overwhelming majority of them have never been subjected to empirical scrutiny. In addition, an ever-increasing contingent of self-help "gurus" on television and radio talk shows routinely offer advice of questionable scientific validity to a receptive, but often vulnerable, audience of troubled individuals (Lilienfeld, 2012).

Similarly questionable practices can be found in the domains of

psychological assessment and diagnosis. Despite well-replicated evidence that statistical (actuarial) formulas are superior to clinical judgment for a broad range of judgmental and predictive tasks (Grove, Zald, Lebow, Snitz, & Nelson, 2000), most clinicians continue to rely on clinical judgment even in cases in which it has been shown to be ill advised (Vrieze & Grove, 2009). There is also evidence that many practitioners tend to be overconfident in their judgments and predictions, and to fall prey to basic errors in reasoning (e.g., confirmation bias, illusory correlation, hindsight bias) in the process of case formulation (see Garb & Boyle, Chapter 2, this volume). Moreover, many practitioners base their interpretations on assessment instruments (e.g., human figure drawing tests, Rorschach Inkblot Test, Myers–Briggs Type Indicator, anatomically detailed dolls) that are either highly controversial or questionable from a scientific standpoint (see Hunsley, Lee, Wood, & Taylor, Chapter 3, this volume).

Still other clinicians render confident diagnoses of psychiatric conditions, such as dissociative identity disorder (known formerly as multiple personality disorder), whose validity remains in dispute (see Lilienfeld & Lynn, Chapter 5, this volume, but see also Gleaves, May, & Cardena, 2001; Reinders, Willemsen, Vos, den Boer, & Nijenhuis, 2012, for different perspectives). The problem of questionable diagnostic labels is especially acute in courtroom settings, where psychiatric labels of unknown or doubtful validity (e.g., road rage syndrome, sexual addiction, battered woman's syndrome) are sometimes invoked as exculpatory defenses (see McCann, Lynn, Lilienfeld, Shindler, & Hammond, Chapter 4, this volume).

STRIKING A BALANCE BETWEEN EXCESSIVE OPEN-MINDEDNESS AND EXCESSIVE SKEPTICISM

Still, we should avoid the temptation to be dismissive. At least some of the largely or entirely untested psychotherapeutic, assessment, and diagnostic methods reviewed in this volume may ultimately prove to be efficacious or valid. It would be a serious error to refuse to consider any untested techniques out of hand or antecedent to prior critical scrutiny. In fairness, such closed-mindedness has sometimes characterized debates concerning the efficacy of novel psychotherapies (Beutler & Harwood, 2001). Nevertheless, a basic tenet of science is that the burden of proof falls squarely on the claimant, not the critic (Shermer, 1997). As a consequence, it is up to the proponents of these techniques to demonstrate that they work, not up to the critics of these techniques to demonstrate the converse.

As Carl Sagan (1995b) eloquently pointed out, scientific inquiry demands a unique mix of open-mindedness and penetrating skepticism (see also Shermer, 2001). We must remain open to novel and untested claims, regardless of how superficially implausible they might appear at first blush. At the same time, we must subject these claims to incisive scrutiny to ensure that they withstand the crucible of rigorous scientific testing. As space

scientist James Oberg observed, keeping an open mind is a virtue but this mind cannot be so open that one's brains fall out (Sagan, 1995a; see also Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume). Although the requirement to hold all claims to high levels of skeptical scrutiny applies to all domains of science, such scrutiny is especially crucial in applied areas, such as clinical psychology, in which erroneous claims or ineffective practices have the potential to produce harm (Lilienfeld, 2007).

WHY POTENTIALLY PSEUDOSCIENTIFIC TECHNIQUES CAN BE HARMFUL

Some might respond to our arguments by contending that although many of the techniques reviewed in this book are either untested or ineffective, most are likely to prove either efficacious or innocuous. From this perspective, our emphasis on the dangers posed by such techniques is misplaced because unresearched mental health practices are at worst inert.

Nevertheless, this counterargument overlooks several important considerations. Specifically, there are at least three major ways in which unsubstantiated mental health techniques can be problematic (Lilienfeld, 2002; see also Beyerstein, 2001). First, some of these techniques may be harmful per se (Dimidjian & Hollon, 2010; Lilienfeld, 2007). The tragic case of Candace Newmaker, the 10-year-old Colorado girl who was smothered to death in 2000 by therapists practicing a variant of rebirthing therapy (see Mercer, Chapter 15, this volume), attests to the dangers of implementing untested therapeutic techniques. There is also increasing reason to suspect that certain suggestive techniques (e.g., hypnosis, guided imagery) for unearthing purportedly repressed memories of childhood trauma may exacerbate or even produce psychopathology by inadvertently implanting false memories of past events (see Pignotti & Thyer, Chapter 7, and Lynn, Krackow, Loftus, Locke, & Lilienfeld, Chapter 8, this volume). Even the use of facilitated communication for infantile autism (see Romanczyk, Turner, Sevlever, & Gillis, Chapter 14, this volume) has resulted in numerous erroneous accusations of child abuse against family members. Moreover, there is accumulating evidence that certain widely used treatment techniques, such as critical incident stress debriefing (see Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume), Scared Straight programs for delinquency (see Petrosino, MacDougall, Hollis-Peel, Fronius, & Guckenberg, Chapter 16, this volume) and perhaps certain self-help programs (Rosen, 1987; see Chapter 9, this volume) can be harmful for some clients. Consequently, the oft-held assumption that "doing something is always better than doing nothing" in the domain of psychotherapy is likely to be mistaken. As psychologist Richard Gist reminds us, doing something is not license to do anything.

Second, even psychotherapies that are by themselves innocuous can indirectly produce harm by depriving individuals of scarce time, financial

resources, or both. Economists refer to this side effect as "opportunity cost." As a result of opportunity cost, individuals who would otherwise use their time and money to seek out demonstrably efficacious treatments may be left with precious little of either. Such individuals may therefore be less likely to obtain interventions that could prove beneficial.

Third, the use of unsubstantiated techniques eats away at the scientific foundations of the profession of clinical psychology (Baker et al., 2008; Lilienfeld, 1998; McFall, 1991). As one of us (Lilienfeld, 2002) observed:

> Once we abdicate our responsibility to uphold high scientific standards in administering treatments, our scientific credibility and influence are badly damaged. Moreover, by continuing to ignore the imminent dangers posed by questionable mental health techniques, we send an implicit message to our students that we are not deeply committed to anchoring our discipline in scientific evidence or to combating potentially unscientific practices. Our students will most likely follow in our footsteps and continue to turn a blind eye to the widening gap between scientist and practitioner, and between research evidence and clinical work. (p. 9)

In addition, the promulgation of treatment and assessment techniques of questionable validity can undermine the general public's faith in the profession of clinical psychology and lead citizens to place less trust in the assertions of clinical researchers and practitioners (Lilienfeld, 2012).

THE DIFFERENCES BETWEEN SCIENCE AND PSEUDOSCIENCE: A PRIMER

One of the major goals of this book is to distinguish scientific from pseudoscientific claims in clinical psychology. To accomplish this goal, however, we must first delineate the principal differences between scientific and pseudoscientific research programs. As one of us has noted elsewhere (Lilienfeld, 1998), science probably differs from pseudoscience in degree rather than in kind. Science and pseudoscience can be thought of as Roschian (Rosch, 1973) or open (Meehl & Golden, 1982; Pap, 1953) concepts that possess intrinsically fuzzy boundaries and an indefinitely extendable list of indicators. Nevertheless, the fuzziness of such categories does not mean that distinctions between science and pseudoscience are fictional or entirely arbitrary. As psychophysicist S. S. Stevens observed, the fact that the precise boundary between day and night is indistinct does not imply that day and night cannot be meaningfully differentiated (see Leahey & Leahey, 1983). From this perspective, pseudosciences can be conceptualized as exhibiting a fallible, but nevertheless useful, list of indicators or "warning signs." The more such warning signs a discipline exhibits, the more it begins to cross the murky dividing line separating science from pseudoscience (see also Herbert et al., 2000). A number of philosophers

of science (e.g., Bunge, 1984) and psychologists (e.g., Ruscio, 2001) have outlined some of the most frequent features of pseudoscience. Among these features are the following (for further discussions, see Herbert et al., 2000; Hines, 1988; Lilienfeld, 1998):

1. *An overuse of ad hoc hypotheses designed to immunize claims from falsification.* From a Popperian or neo-Popperian standpoint (see Popper, 1959), assertions that could never in principle be falsified are unscientific (but see McNally, 2003, for a critique of Popperian notions). The repeated invocation of ad hoc hypotheses to explain away negative findings is a common tactic among proponents of pseudoscientific claims. Moreover, in most pseudosciences, ad hoc hypotheses are simply "pasted on" to plug holes in the theory in question. When taken to an extreme, ad hoc hypotheses can provide an impenetrable barrier against potential refutation. For example, some proponents of eye movement desensitization and reprocessing (EMDR) have argued that negative findings concerning EMDR are almost certainly attributable to low levels of fidelity to the treatment procedure (see Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume). But they have typically been inconsistent in their application of the treatment fidelity concept (Rosen, 1999).

It is crucial to emphasize that the invocation of ad hoc hypotheses in the face of negative evidence is sometimes a legitimate strategy in science. In scientific research programs, however, such maneuvers tend to enhance the theory's content, predictive power, or both (see Lakatos, 1978; Meehl, 1990).

2. *Absence of self-correction.* Scientific research programs are not necessarily distinguished from pseudoscientific research programs in the verisimilitude of their claims because proponents of both programs frequently advance incorrect assertions. Nevertheless, in the long run most scientific research programs tend to eliminate these errors, whereas most pseudoscientific research programs do not. Consequently, intellectual stagnation is a hallmark of most pseudoscientific research programs (Ruscio, 2001). For example, astrology has changed remarkably little in the past 2,500 years (Hines, 1988).

3. *Evasion of peer review.* On a related note, many proponents of pseudoscience avoid subjecting their work to the often ego-bruising process of peer review (Ruscio, 2001; see also Gardner, 1957, for illustrations). In some cases, they may do so on the grounds that the peer review process is inherently biased against findings or claims that contradict well-established paradigms (e.g., see Callahan, 2001a, for an illustration involving Thought Field Therapy; see also Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume). In other cases, they may avoid the peer review process on the grounds that their assertions cannot be evaluated adequately using standard scientific methods. Although the peer review process is far from flawless

(see Peters & Ceci, 1982, for a striking example), it remains the best mechanism for self-correction in science and assists investigators in identifying errors in their reasoning, methodology, and analyses. By remaining largely insulated from the peer review process, some proponents of pseudoscience forfeit an invaluable opportunity to obtain corrective feedback from informed colleagues.

4. *Emphasis on confirmation rather refutation.* The brilliant physicist Richard Feynman (1985) maintained that the essence of science is a bending over backwards to prove oneself wrong. Bartley (1962) similarly maintained that science at its best involves the maximization of constructive criticism. Ideally, scientists subject their cherished claims to grave risk of refutation (Meehl, 1978; see also Ruscio, 2001). In contrast, pseudoscientists tend to seek only confirming evidence for their claims. Because a determined advocate can find at least some supportive evidence for virtually any claim (Popper, 1959), this confirmatory hypothesis-testing strategy is not an efficient means of rooting out error in one's web of beliefs.

Moreover, as Bunge (1967) observed, most pseudosciences manage to reinterpret negative or anomalous findings as corroborations of their claims (see Herbert et al., 2000). For example, proponents of extrasensory perception (ESP) have sometimes interpreted isolated cases of *worse* than chance performance on parapsychological tasks (known as "psi missing") as evidence of ESP (Gilovich, 1991; Hines, 1988).

5. *Reversed burden of proof.* As noted earlier, the burden of proof in science rests on the individual making a claim, not on the critic. Proponents of pseudoscience frequently flout this principle and instead demand that skeptics demonstrate beyond a reasonable doubt that a claim (e.g., an assertion regarding the efficacy of a novel therapeutic technique) is false. This error is similar to the logician's ad ignorantium fallacy (i.e., the argument from ignorance)—the mistake of assuming that a claim is likely to be correct merely because there is no compelling evidence against it (Shermer, 1997). For example, some proponents of unidentified flying objects (UFOs) have insisted that skeptics account for every unexplained report of an anomalous event in the sky (Hines, 1988; Sagan, 1995a). But because it is essentially impossible to prove a universal negative, this tactic incorrectly places the burden of proof on the skeptic rather than the claimant.

6. *Absence of connectivity.* In contrast to most scientific research programs, pseudoscientific research programs tend to lack "connectivity" with other scientific disciplines (Bunge, 1983; Stanovich, 2012). In other words, pseudosciences often purport to create entirely new paradigms out of whole cloth rather than to build on extant paradigms. In so doing, they often neglect well-established scientific principles or hard-won scientific knowledge. For example, many proponents of ESP argue that it is a genuine (although heretofore undetected) physical process of perception, even though reported cases of ESP violate almost every major law of physical signals (e.g., ESP purportedly operates just as strongly from thousands of

miles away as it does from a few feet away). Although scientists should remain open to the possibility that an entirely novel paradigm has successfully overturned all preexisting paradigms, they must insist on extremely high standards of evidence before drawing such a conclusion. This dictum comports with Bayesian perspectives on science, which mandate that *a priori* plausibility be considered when evaluating the likelihood of scientific theories (Lilienfeld, 2011; Wagenmakers, Wetzels, Boorsbom, & van der Maas, 2011).

7. *Overreliance on testimonial and anecdotal evidence.* Testimonial and anecdotal evidence can be quite useful in the early stages of scientific investigation. Nevertheless, such evidence is almost always much more helpful in the context of discovery (i.e., hypothesis generation) than in the context of justification (i.e., hypothesis testing; see Reichenbach, 1938). Proponents of pseudoscientific claims frequently invoke reports from selected cases (e.g., "This treatment clearly worked for Person X, because Person X improved markedly following the treatment") as a means of furnishing dispositive evidence for these claims. For example, proponents of certain treatments (e.g., secretin, glutein-free diets, chelation therapy) for autistic spectrum disorder (see Waschbusch & Waxmonsky, Chapter 13, this volume) have often pointed to uncontrolled case reports of improvement as supportive evidence (Offit, 2010).

As Gilovich (1991) observed, however, case reports almost never provide sufficient evidence for a claim, although they often provide necessary evidence for this claim. For example, if a new form of psychotherapy is efficacious, one should certainly expect at least some positive case reports of improvement. But such case reports do not provide adequate evidence that the improvement was attributable to the psychotherapy because this improvement could have been produced by a host of other influences (e.g., placebo effects, regression to the mean, spontaneous remission, or maturation; see Cook & Campbell, 1979; Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013).

8. *Use of obscurantist language.* Many proponents of pseudoscience use impressive sounding or highly technical jargon in an effort to provide their disciplines with the superficial trappings of science (see van Rillaer, 1991, for a discussion of "strategies of dissimulation" in pseudoscience). Such language may be convincing to individuals unfamiliar with the scientific underpinnings of the claims in question and may therefore lend these claims an unwarranted imprimatur of scientific legitimacy.

For example, the developer of EMDR explained the efficacy of this treatment as follows (see also Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume):

> [The] valences of the neural receptors (synaptic potential) of the respective neuro networks, which separately store various information plateaus and levels of adaptive information, are represented by the letters Z

> through A. It is hypothesized that the high-valence target network (Z) cannot link up with the more adaptive information, which is stored in networks with a lower valence. That is, the synaptic potential is different for each level of affect held in the various neuro networks. . . . The theory is that when the processing system is catalyzed in EMDR, the valence of the receptors is shifted downward so that they are capable of linking with the receptors of the neuro networks with progressively lower valences. (Shapiro, 1995, pp. 317–318)

9. *Absence of boundary conditions.* Most well-supported scientific theories possess boundary conditions, that is, well-articulated limits under which predicted phenomena do and do not apply. In contrast, many or most pseudoscientific phenomena are purported to operate across an exceedingly wide range of conditions. As Hines (1988, 2001) noted, one frequent characteristic of fringe psychotherapies is that they are ostensibly efficacious for almost all disorders regardless of their etiology. For example, some proponents of Thought Field Therapy (see Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume) have proposed that this treatment is beneficial for virtually all mental disorders. Moreover, the developer of this treatment has posited that it is efficacious not only for adults but for "horses, dogs, cats, infants, and very young children" as well (Callahan, 2001b, p. 1255).

10. *The mantra of holism.* Proponents of pseudoscientific claims, especially in organic medicine and mental health, often resort to the "mantra of holism" (Ruscio, 2001) to explain away negative findings. When invoking this mantra, they typically maintain that scientific claims can be evaluated only within the context of broader claims and therefore cannot be judged in isolation. For example, some proponents of the Rorschach Inkblot Test have responded to criticisms of this technique (see Hunsley, Lee, Wood, & Taylor, Chapter 3, this volume) by asserting that clinicians virtually never interpret results from a Rorschach in isolation. Instead, in actual practice clinicians consider numerous pieces of information, only one of which may be a Rorschach protocol. There are two major difficulties with this line of reasoning. First, it implies that clinicians can effectively integrate in their heads a great deal of complex psychometric information from diverse sources, a claim that is doubtful given the research literature on clinical judgment (see Garb & Boyle, Chapter 2, this volume). Second, by invoking the mantra of holism, proponents of the Rorschach and other techniques can readily avoid subjecting their claims to the risk of falsification. In other words, if research findings corroborate the validity of a specific Rorschach index, Rorschach proponents can point to these findings as supportive evidence, but if these findings are negative, Rorschach proponents can explain them away by maintaining that "clinicians never interpret this index in isolation anyway" (see Merlo & Barnett, 2001, for an example). This "heads I win, tails you lose" reasoning places the claims of these proponents largely outside of the boundaries of science.

We encourage readers to bear in mind the aforementioned list of pseudoscience indicators (see Lilienfeld, Ammirati, & David, 2012, and Ruscio, 2001, for other useful indicators) when evaluating the claims presented in this volume. At the same time, we remind readers that these indicators are only probabilistically linked to pseudoscientific research programs. Scientists, even those who are well trained, are not immune from such practices. In scientific research programs, however, such practices tend eventually to be weeded out through the slow but steady process of self-correction. In contrast to sciences, in which erroneous claims tend to be gradually ferreted out by a process akin to natural selection (e.g., see Campbell's [1974] discussion of evolutionary epistemology), pseudosciences tend to remain stagnant in the face of contradictory evidence.

CONSTRUCTIVE EFFORTS TO ADDRESS THE PROBLEM

Until recently, the field of clinical psychology has shown relatively little interest in addressing the threats posed by pseudoscientific or otherwise questionable practices. As Paul Meehl (1993), perhaps the foremost clinical psychologist of the latter half of the 20th century, observed:

> It is absurd, as well as arrogant, to pretend that acquiring a Ph.D. somehow immunizes me from the errors of sampling, perception, recording, retention, retrieval, and inference to which the human mind is subject. In earlier times, all introductory psychology courses devoted a lecture or two to the classic studies in the psychology of testimony, and one mark of a psychologist was hard-nosed skepticism about folk beliefs. It seems that quite a few clinical psychologists never got exposed to this basic feature of critical thinking. My teachers at [the University of] Minnesota . . . shared what Bertrand Russell called the dominant passion of the true scientist—the passion not to be fooled and not to fool anybody else . . . all of them asked the two searching questions of positivism: "What do you mean?" "How do you know?" If we clinicians lose that passion and forget those questions, we are little more than be-doctored, well-paid soothsayers. I see disturbing signs that this is happening and I predict that, if we do not clean up our clinical act and provide our students with role models of scientific thinking, outsiders will do it for us. (pp. 728–729)

Nevertheless, the past two decades have witnessed several constructive efforts to address the problems posed by questionable and potentially pseudoscientific methods in clinical psychology. In particular, Division 12 of the American Psychological Association has advanced a set of criteria for empirically supported treatments (ESTs) for adult and childhood disorders, along with provisional lists of therapeutic techniques that satisfy these criteria (see Chambless & Ollendick, 2001, and Barlow, 2004, for

thoughtful reviews). Vigorous and healthy debate surrounds the criteria established for identifying ESTs as well as the current list of ESTs (Herbert, 2003; Westen, Novotny, & Thompson-Brenner, 2004; see also Gaudiano, Dalrymple, Weinstock, & Lohr, Chapter 6, this volume). Despite this controversy, it seems clear that the increasing push toward ESTs reflects a heightened emphasis on distinguishing interventions that are scientifically supported from those whose support is negligible or nonexistent. In this respect, the EST movement, although hardly immune from criticism (see Lilienfeld, Lynn, & Lohr, Chapter 17, this volume), is an important step in the direction of minimizing error in clinical inference.

This and other developments which reflect a heightened emphasis on evidence-based practice in some doctoral training programs (Weissman et al., 2006), suggest that careful attention is at long last being accorded to questionable practices in clinical psychology and to distinguishing them from practices with stronger evidentiary support. We hope that readers will find this second edition of this edited volume to represent another constructive step in this direction.

THE GOALS OF THIS VOLUME

With the aforementioned considerations in mind, the primary goal of this second edition of this edited volume is to assist readers—whom we hope will include clinical researchers, practicing psychologists, psychiatrists, social workers, counselors, and psychiatric nurses, graduate students in clinical psychology and allied disciplines (e.g., social work, counseling), medical students, lawyers, educators, and educated laypersons—with the crucial task of distinguishing techniques in clinical psychology that are scientifically supported or promising from those that are scientifically unsupported or untested. To assist readers with this task, we have asked the authors of each chapter to delineate not only which techniques and claims are devoid of empirical support, but also which are either empirically supported or promising. In this way, we expect readers to emerge with an enhanced understanding and appreciation of the differences between mental health techniques that are and are not grounded in the most up-to-date scientific evidence. In addition, as noted earlier, we intend to assist readers with the task of identifying research programs in clinical psychology that embody many of the features of pseudoscience and to distinguish them from research programs that exemplify the core features of scientific epistemology (e.g., self-correction).

We have organized this volume into four major sections. First, we begin with an examination of questionable or untested practices and assumptions in the domains of psychological assessment and diagnosis. Second, we examine general controversies in psychotherapy and self-help interventions that cut across multiple psychological disorders. Third, we turn to

largely untested or unsubstantiated treatment techniques (both psychotherapeutic and psychopharmacological) for various adult psychological conditions, including posttraumatic stress disorder, alcoholism, and depression. Fourth, we examine similarly untested and unsubstantiated treatments for childhood disorders, with a particular focus on attention-deficit/hyperactivity disorder and infantile autism. We conclude the volume with a brief set of constructive remedies for narrowing the gap between scientist and practitioner.

By concluding this volume on a relatively optimistic note, we intend to leave readers with the impression that the problem of pseudoscience in contemporary clinical psychology, though formidable in severity and scope, may not be intractable. If our sanguine assessment is correct, a future generation of clinical psychologists may perceive this volume as a mere historical curiosity, a legacy of a bygone era when clinical practices were often unsubstantiated and not routinely grounded in the best available scientific evidence. Nothing would please us more.

REFERENCES

Baker, T. B., McFall, R. M., & Shoham, V. (2008). Current status and future prospects of clinical psychology toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest*, *9*, 67–103.

Barlow, D. H. (2004). Psychological treatments. *American Psychologist*, *59*, 869–878.

Bartley, W. W. (1962). *The retreat to commitment*. New York: Knopf.

Beutler, L. E., & Harwood, T. M. (2001). Antiscientific attitudes: What happens when scientists are unscientific? *Journal of Clinical Psychology, 57*, 43–51.

Beyerstein, B. L. (2001). Fringe psychotherapies: The public at risk. *Scientific Review of Alternative Medicine, 5*, 70–79.

Bunge, M. (1967). *Scientific research*. New York: Springer.

Bunge, M. (1983). Speculation: Wild and sound. *New Ideas in Psychology, 1*, 3–6.

Bunge, M. (1984, Fall). What is pseudoscience? *Skeptical Inquirer, 9*, 36–46.

Callahan, R. J. (2001a). The impact of Thought Field Therapy on heart rate variability. *Journal of Clinical Psychology, 57*, 1154–1170.

Callahan, R. J. (2001b). Thought Field Therapy: Response to our critics and a scrutiny of some old ideas of science. *Journal of Clinical Psychology, 57*, 1251–1260.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl R. Popper* (pp. 412–463). LaSalle, IL: Open Court.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685–716.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Corsini, R. J. (Ed.). (2001). *Handbook of innovative therapy* (2nd ed.). New York: Wiley.

Dawes, R. M. (1994). *House of cards*: *Psychology and psychotherapy built on myth*. New York: Free Press.

Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist*, *65*, 21–33.

Eisner, D. A. (2000). *The death of psychotherapy: From Freud to alien abductions*. Westport, CT: Praeger.

Feinstein, D., Eden, D., Craig, G., & Bowen, M. (2005). *The promise of energy psychology*: *Revolutionary tools for dramatic personal change*. London: Penguin Books.

Feynman, R. P. (with R. Leighton). (1985). *Surely you're joking, Mr. Feynman*: *Adventures of a curious character*. New York: Norton.

Fox, R. E. (1996). Charlatanism, scientism, and psychology's social contract. *American Psychologist, 51*, 777–784.

Gambrill, E. (2006). *Critical thinking in clinical practice*: *Improving the quality of judgments and decisions*. New York: Wiley.

Gardner, M. (1957). *Fads and fallacies in the name of science*. New York: Dover.

Gilovich, T. (1991). *How we know what isn't so*: *The fallibility of human reason in everyday life*. New York: Free Press.

Gleaves, D. H., May, M. C., & Cardena, E. (2001). An examination of the diagnostic validity of dissociative identity disorder. *Clinical Psychology Review, 21*, 577–608.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.

Heller, M. C., & Duclos, M. (2012). *Body psychotherapy*: *History, concepts, and methods*. New York: Norton.

Herbert, J. D. (2003). Defining empirically supported treatments: Pitfalls and possible solutions. *Behavior Therapist, 23*, 113–134.

Herbert, J. D., Lilienfeld, S. O., Lohr, J. M., Montgomery, R. W., O'Donohue, W. T, Rosen, G. M., et al. (2000). Science and pseudoscience in the development of eye movement desensitization and reprocessing. *Clinical Psychology Review, 20*, 945–971.

Hines, T. (1988). *Pseudoscience and the paranormal*: *A critical examination of the evidence*. Buffalo, NY: Prometheus Books.

Hines, T. M. (2001). The Doman–Delcato patterning treatment for brain damage. *Scientific Review of Alternative Medicine, 5*, 80–89.

Kalal, D. M. (1999, April). Critical thinking in clinical practice: Pseudoscience, fad psychology, and the behavior therapist. *Behavior Therapist*, 81–84.

Lakatos, I. (1978). *Philosophical papers: Vol. 1. The methodology of scientific research programmes* (J. Worrall & G. Currie, Eds.). New York: Cambridge University Press.

Leahey, T. H., & Leahey, G. E. (1983). *Psychology's occult doubles*: *Psychology and the problem of pseudoscience*. Chicago: Nelson–Hall.

Lilienfeld, S. O. (1998). Pseudoscience in contemporary clinical psychology: What it is and what we can do about it. *Clinical Psychologist, 51*, 3–9.

Lilienfeld, S. O. (2002). The scientific review of mental health practice: Our raison d'être. *Scientific Review of Mental Health Practice, 1*, 5–10.
Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*, 53–70.
Lilienfeld, S. O. (2011). Distinguishing scientific from pseudoscientific psychotherapies: Evaluating the role of theoretical plausibility, with a little help from Reverend Bayes. *Clinical Psychology: Science and Practice*, *18*, 105–112.
Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist*, *67*, 111–129.
Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *50*, 7–36.
Lilienfeld, S. O., & Lohr, J. M. (2000). News and comment: Thought Field Therapy educators and practitioners sanctioned. *Skeptical Inquirer, 24*, 5.
Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2003). *Science and pseudoscience in clinical psychology*. New York: Guilford Press.
Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review*, *33*, 883–900.
McFall, R. M. (1991). Manifesto for a science of clinical psychology. *Clinical Psychologist, 44*, 75–88.
McNally, R. J. (2003). The demise of pseudoscience. *Scientific Review of Mental Health Practice*, *2*(2), 97–101.
Meehl, P. E. (1978). Theoretical risks or tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 816–834.
Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*, 108–141.
Meehl, P. E. (1993). Philosophy of science: Help or hindrance? *Psychological Reports, 72*, 707–733.
Meehl, P. E., & Golden, R. R. (1982). Taxometric methods. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–181). New York: Wiley.
Mercer, J. (2002). Attachment therapy: A treatment without empirical support. *Scientific Review of Mental Health Practice*, *1*, 9–16.
Merlo, L., & Barnett, D. (2001, September). All about inkblots. *Scientific American, 285*, 13.
Offit, P. A. (2008). *Autism's false prophets*. New York: Columbia University Press.
Pap, A. (1953). Reduction sentences and open concepts. *Methodos, 5*, 3–30.
Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5*, 187–255.
Pignotti, M., & Thyer, B. A. (2009). Use of novel unsupported and empirically supported therapies by licensed clinical social workers: An exploratory study. *Social Work Research*, *33*, 5–17.
Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Illinois Press.

Reinders, A. S., Willemsen, A. T., Vos, H. P., den Boer, J. A., & Nijenhuis, E. R. (2012). Fact or factitious?: A psychobiological study of authentic and simulated dissociative identity states. *PLOS ONE*, *7*(6), e39279.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.

Rosen, G. M. (1987). Self-help treatment books and the commercialization of psychotherapy. *American Psychologist, 42*, 46–51.

Rosen, G. M. (1999). Treatment fidelity and research on eye movement desensitization and reprocessing. *Journal of Anxiety Disorders, 13*, 173–184.

Ruscio, J. (2001). *Clear thinking with psychology*: *Separating sense from nonsense*. Pacific Grove, CA: Wadsworth.

Sagan, C. (1995a). *The demon-haunted world*: *Science as a candle in the dark*. New York: Random House.

Sagan, C. (1995b, January/February). Wonder and skepticism. *Skeptical Inquirer, 19*, 24–30.

Shannon, S. (Ed.). (2002). *Handbook of complementary and alternative therapies in mental health*. San Diego, CA: Academic Press.

Shapiro, F. (1995). *Eye movement desensitization and reprocessing*: *Basic protocols, principles, and procedures*. New York: Guilford Press.

Shermer, M. (1997). *Why people believe weird things*: *Pseudoscience, superstition, and other confusions of our time*. New York: Freeman.

Shermer, M. (2001). *The borderlands of science*: *Where sense meets nonsense*. New York: Oxford University Press.

Singer, M. T., & Lalich, J. (1996). *Crazy therapies*: *What are they? Do they work*? San Francisco: Jossey-Bass.

Stanovich, K. (2012). *How to think straight about psychology* (10th ed.). New York: HarperCollins.

van Rillaer, J. (1991). Strategies of dissimulation in the pseudosciences. *New Ideas in Psychology, 9*, 235–244.

Vrieze, S. I., & Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psychology*: *Research and Practice*, *40*, 525–531.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*, 426–432.

Weissman, M. M., Verdeli, H., Gameroff, M. J., Bledsoe, S. E., Betts, K., Mufson, L., et al. (2006). National survey of psychotherapy training in psychiatry, psychology, and social work. *Archives of General Psychiatry*, *63*, 925–934.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, *130*, 631–663.

PART I

CONTROVERSIES IN ASSESSMENT AND DIAGNOSIS

CHAPTER TWO

Understanding Why Some Clinicians Use Pseudoscientific Methods

Findings from Research on Clinical Judgment

Howard N. Garb and Patricia A. Boyle

If pseudoscientific methods are not valid, then why are grand claims made for them and why do some clinicians think that they are valid? There are many reasons, and some reasons may apply to some clinicians but not to others. Obviously, if one develops a new assessment instrument or treatment intervention, there are personal and financial reasons for believing in, and overstating, the value of one's product. Once grand claims have been made, clinicians may use an assessment instrument or treatment intervention to see if it works. However, once clinicians use pseudoscientific methods with clients, why do they sometimes not see that they are invalid and ineffective?

The purpose of this chapter is to explain why it can be difficult to learn from clinical experience. Research on the value of clinical experience and clinical training will be reviewed along with research on cognitive processes. Comments will be made about the nature of feedback that clinicians receive in clinical settings. Finally, recommendations will be made for improving clinical practice.

THE VALUE OF CLINICAL EXPERIENCE AND TRAINING

Clinical lore suggests that psychologists and mental health professionals learn from experience by working with clients in clinical settings.

Experienced clinicians are presumed to make more accurate and valid assessments of personality and psychopathology than less experienced graduate students and mental health providers. Similarly, presumed experts are assumed to be more competent providers of psychological interventions than other clinicians. Psychology training programs adhere to these assumptions, and common supervisory practices emphasize the value of experience in the development of competent clinicians. The inherent message to mental health trainees is that clinical acumen develops over time and with increased exposure to various clients and presenting problems.

Although research has often supported the value of training, learning from clinical experience has had less support. Narrative reviews of clinical judgment have concluded that when clinicians are given identical sets of information, experienced clinicians are no more accurate than less experienced clinicians and graduate students, though they may be better at structuring judgment tasks (e.g., generating questions during an interview; Dawes, 1994; Garb, 1989, 1998, 2005; Garb & Schramke, 1996; Goldberg, 1968; Tracey, Wampold, Lichtenberg, & Goodyear, 2014; Wiggins, 1973; see also Meehl, 1997). Similarly, a recent meta-analysis (Spengler et al., 2009) found only a small positive effect for training and experience.

Results that reflect the difficulty of learning from experience are not unique to mental health settings. For example, in an article in the *Annals of Internal Medicine*, Choudhry, Fletcher, and Soumerai (2005, p. 260) concluded that "physicians who have been in practice longer may be at risk for providing lower-quality care."

Results from a meta-analysis will be described, followed by a narrative review of the literature on the value of clinical experience and training.

Meta-Analysis on Experience and Validity of Judgments

Meta-analyses use statistical techniques to describe results across studies. They can boost statistical power, allowing for the identification of small but statistically significant differences that would not be detected by analyzing data from only a single study.

A meta-analysis has been conducted on experience, training, and clinical judgment in the area of mental health (Spengler et al., 2009). The authors synthesized results from 75 clinical judgment studies. A finding they emphasized is that the combined effect of training and experience was small but positive (d = 0.12; this is equivalent to a correlation of about r = 0.06). Some of their other results differed more sharply from the narrative reviews. Notably, the effects of training and experience were not significantly different. Also, Spengler et al. concluded that having specific training and experience with a judgment task was unrelated to validity. Their conclusion was intended to be broad and to cover all mental health groups that receive specific training (e.g., those who have received training

in neuropsychology, forensic psychology, behavioral sleep medicine, or gerontology).

Is it true that there is no significant difference between the effect of training and the effect of experience on the validity of judgments? And that the effect of specific training and experience, including the effect of training in neuropsychological assessment, is unrelated to validity? To understand the results of a meta-analysis, one needs to be familiar with the range of studies that have been included. Spengler et al. used more inclusive criteria than those used in the narrative reviews cited above. For example, to learn about the effects of training and experience on validity, they included not only studies on the validity of judgments, but also studies on cognitive processes (e.g., speed of making a judgment, cognitive heuristics). Cognitive heuristics, along with cognitive biases, will be described later in this chapter, but here it is noted that to evaluate the effect of training and experience on the validity of judgments, one must directly study the validity of judgments rather than the cognitive processes of clinicians (see Funder, 1987). Spengler et al. also included judgment tasks related to vocational issues and career counseling. With regard to judges, they included pastoral counselors and psychiatric nurses. Finally, some of the tasks in the studies may not seem relevant to current clinical practice, for example, being able to decipher whether a dream was reported by a patient with or without schizophrenia (Brenneis, 1971).

To evaluate the claim made by Spengler et al. that the effects of specific training and specific experiences are unrelated to validity, one can turn to other bodies of research. For example, a series of studies provides evidence that training in neuropsychology leads to an increase in the validity of judgments. Many clinical psychologists screen for brain impairment (e.g., using the Bender Visual Motor Gestalt Test), and results from six studies indicate that their hit rates vary from 53 to 73% (Garb, 1998, p. 158). For neuropsychologists using a neuropsychological test battery, their average hit rate across 11 studies was 84% (Garb & Schramke, 1996).

As a general rule, a meta-analysis will be valuable when a substantial number of studies have been conducted. As the topic for a meta-analysis becomes broader and more complex, more studies are needed. An especially large number of studies are needed for a meta-analysis that attempts to cover all mental health judgment tasks using all types of assessment information across all mental health professional fields. And if the meta-analysis also includes research on cognitive processes and biases, then even more studies will be needed.

The complexity of this area provides a challenge for meta-analyses. For example, to evaluate the effect of training, one would want to (1) compare lay judges or undergraduates to graduate students, interns, or clinicians, (2) follow interns or graduate students over time, and (3) evaluate the effects of specialized training provided to clinicians, interns, graduate students, or lay judges. Interactions may occur; for example, training may

have a positive effect when using one type of assessment information but not another for some types of judgment tasks but not for others.

Given that there may not be a sufficient density of studies for a meta-analysis on such a broadly defined and complex topic, a narrative review on the value of training and experience appears next. In this narrative review, validity is defined narrowly: Studies on reliability and bias are not included. Results are described for: (1) experienced versus less experienced clinicians, (2) clinicians versus trainees, (3) clinicians versus graduate students, (4) advanced graduate students versus beginning graduate students, (5) graduate students followed over time, (6) clinicians and graduate students versus lay judges, and (7) clinicians differing in experience and specialized training.

Experienced versus Less Experienced Clinicians

For the task of interpreting personality assessment test results, alleged experts have not been more accurate than other clinicians, and experienced clinicians have not been more accurate than less experienced clinicians (Graham, 1967; Levenberg, 1975; Turner, 1966; Walters, White, & Greene, 1988; Wanderer, 1969; Watson, 1967). For example, Graham (1967) presented Minnesota Multiphasic Personality Inventory (MMPI) protocols to two groups of psychologists with differing levels of experience with the test. The first group consisted of PhD-level psychologists who had used the MMPI routinely in clinical practice for approximately 5 years. The second group consisted of PhD-level psychologists who had used the MMPI much more frequently in clinical practice for over 5 years and who also demonstrated a broad knowledge of the research literature on the MMPI. Clinicians were asked to perform Q-sorts to describe patients' personality features, and clinician ratings were compared with criterion Q-sorts that were generated on the basis of patient and family member interviews. Although results indicated that overall judgmental validity was moderate, with correlations ranging from .29 to .37, judgmental validity was *unrelated* to experience with the MMPI. That is, clinicians with more MMPI experience and knowledge were no more accurate in their interpretations of MMPI data than were clinicians with significantly less MMPI experience and knowledge.

The relation between experience and validity has also been investigated among psychiatrists. In one study, Kendell (1973) studied the relation between experience and diagnostic accuracy among practicing psychiatrists with varying degrees of experience in the field (each had a minimum of 4 years of practice). Psychiatrists observed segments of initial clinical interviews with psychiatric inpatients and were asked to provide diagnoses on the basis of the interview data. Their diagnoses were compared to diagnoses that were based on more complete information, including full interviews with the patients and the patients' relatives, information from

the case notes of earlier admissions, and additional information collected during patients' stays in the hospital. Results indicated that experience was unrelated to the validity of diagnoses. Similarly, in another study (Muller & Davids, 1999), experienced psychiatrists were no more adept than less experienced clinicians when the task was to assess positive and negative symptoms of schizophrenia.

In a final study using psychiatrists as participants (Hermann, Ettner, Dorwart, Langman-Dorwart, & Kleinman, 1999), number of years of clinical experience was *negatively* related to validity. Hermann et al. reported that "psychiatrists trained in earlier eras were more likely to use ECT [electroconvulsive therapy] for diagnoses outside evidence-based indications" (p. 1059). In this study, older psychiatrists may have made less valid judgments than younger psychiatrists because education regarding the proper and improper uses of ECT has improved in recent years. If this is true, then having years of clinical experience did not compensate for not having up-to-date training.

Similar results have been obtained in the area of neuropsychology: Neuropsychologists with the American Board of Professional Psychology (ABPP) diploma are generally no more accurate than less experienced doctoral-level neuropsychologists (Faust et al., 1988; Gaudette, 1992; Heaton, Smith, Lehmann, & Vogt, 1978; Wedding, 1983). One of the best-known studies was conducted by Faust and colleagues (1988). They examined the validity of judgments made by 155 neuropsychologists. The neuropsychologists evaluated results from several commonly used neuropsychological tools (including the Halstead–Reitan Battery) and made ratings regarding the presence/absence of neurological impairment, as well as ratings of the likely location, process, and etiology of the neurological injury. In addition, the clinicians' levels of training and experience were assessed. Measures of training included cumulative practicum experience in neuropsychology, cumulative supervised neuropsychology hours, relevant coursework, specialized neuropsychology internship training, and the completion of postdoctoral training in neuropsychology. Measures of experience included years of practice in neuropsychology and total career hours spent on issues related to neuropsychology. ABPP status was used as a measure of alleged expertise. Training, experience, and alleged expertise were *not* related to the validity of judgments.

When clinicians have had to structure the judgment task, experience has been related to validity. Brammer (2002) reported that when psychologists were required to request information in a simulated interview, more experienced practitioners made more accurate diagnoses, even controlling for level of training. Thus, more experienced clinicians may be better at knowing what information to collect.

In conclusion, when clinicians are given identical sets of information, experienced clinicians are generally no more accurate than less experienced clinicians. When practitioners are required to search for information or

decide what judgments should be made, experience may be related to validity for some judgment tasks.

Clinicians versus Trainees

Results have been no more promising when clinicians have been compared to trainees. In one study (Hannan et al., 2005; also see Whipple & Lambert, 2011, for additional details), 20 trainees and 20 licensed professionals at a university outpatient clinic were instructed to predict outcomes for clients they were seeing in counseling. In particular, they were instructed to predict which of their clients would be worse off at the end of treatment. Forty of 550 patients deteriorated by the end of treatment (as measured by the Outcome Questionnaire–45 [OQ-45]; Lambert, 2004). Only 3 of the 550 clients had been predicted by their therapist to leave treatment worse off than when they began (one of the three predictions was correct). The experienced therapists did not identify a single client who had deteriorated.

Clinicians versus Graduate Students

One would assume that clinical and counseling psychologists are more accurate than psychology graduate students. However, in empirical studies, clinicians have rarely been more accurate than graduate students, regardless of the type of information provided to clinicians. Studies have revealed no differences in accuracy between experienced clinicians and graduate students when judgments are made on the basis of interview data (Anthony, 1968; Schinka & Sines, 1974), biographical and history information (Oskamp, 1965; Witteman & van den Bercken, 2007), behavioral observation data (Garner & Smith, 1976; Walker & Lewine, 1990), data from therapy sessions (Brenner & Howard, 1976), MMPI protocols (Chandler, 1970; Danet, 1965; Goldberg, 1965, 1968; Graham, 1967, 1971; Oskamp, 1962; Walters et al., 1988; Whitehead, 1985), projective-drawing protocols (Levenberg, 1975; Schaeffer, 1964; Stricker, 1967), Rorschach protocols (Gadol, 1969; Turner, 1966; Whitehead, 1985; see also Hunsley, Lee, Wood, & Taylor, Chapter 3, this volume), screening instruments for detecting neurological impairment (Goldberg, 1959; Leli & Filskov, 1981, 1984; Robiner, 1978), and all of the data that clinical and counseling psychologists usually have available in clinical practice (Johnston & McNeal, 1967).

When statistically significant differences have been found, the results have been mixed. Clinicians did outperform graduate students in two studies, but only when the graduate students were just beginning their training (Falvey & Hebert, 1992; Grebstein, 1963; reanalyzed by Hammond, Hursch, & Todd, 1964). In a third study (Jackson, Rogers, & Shuman, 2004), 134 graduate students and 88 doctoral-level forensic psychologists used comprehensive case materials to predict sexual violence. The authors

were able to examine the validity of the predictions as well as how well legal standards were applied. The graduate students made more accurate predictions of sexual violence, but the forensic psychologists more precisely applied the relevant legal standard.

Advanced Graduate Students versus Beginning Graduate Students

When case history information was given to graduate students, the effect of training on the validity of diagnoses was positive and statistically significant (Morgan, Olson, Krueger, Schellenberg, & Jackson, 2000). Graduate students with more than 15 weeks of training in diagnosis made more valid diagnoses than graduate students with less than 15 weeks of training. When graduate students were taught how to use diagnostic decision trees, there was a statistically significant increase in diagnostic accuracy for all of the graduate students.

Graduate Students Followed over Time

To study the relations between (1) training and validity and (2) experience and validity, it is helpful to conduct longitudinal studies. In one study (Aronson & Akamatsu, 1981), 12 graduate students' judgments were evaluated three times: (1) after the first year of graduate education in psychology, (2) following the completion of a course on MMPI interpretation, and (3) following the completion of a year-long assessment and therapy practicum. Graduate students used the MMPI to make personality ratings. Their ratings were then compared with criterion ratings made on the basis of patient and family interviews. Results revealed a validity coefficient of .20 at time 1, .42 at time 2, and .44 at time 3. Thus, graduate students were able to make more accurate judgments of client profiles following specialized didactic training, but additional practicum experience did not improve accuracy (for related results, see Whitehead, 1985).

Clinicians and Graduate Students versus Lay Judges

Results are mixed on whether clinicians and graduate students are more accurate than lay judges. For example, when instructed to describe psychopathology using interview data, clinicians and graduate students outperformed undergraduate students and individuals with bachelor's degrees (Brammer, 2002; Waxer, 1976). When asked to make similar judgments on the basis of biographical and history data, clinicians outperformed lay judges with respect to the assessment of psychiatric patients (Horowitz, 1962; Lambert & Wertheimer, 1988; Stelmachers & McHugh, 1964; see also Holmes & Howard, 1980), but not with respect to the assessment of normal subjects (Griswold & Dana, 1970; Oskamp, 1965; Weiss, 1963).

Of course, clinicians rarely provide assessments for individuals who are not seeking, or are not currently involved, in treatment. Finally, in a more recent study (Ebling & Levenson, 2003), lay judges (e.g., newlyweds, recently divorced individuals) and professional judges (e.g., marital therapists, marital researchers) watched three-minute videos of 10 married couples. They rated marital satisfaction and predicted whether the marriages would end in divorce. For ratings of marital satisfaction, lay judges were more accurate than professional judges. For predicting divorce, group differences in accuracy were not statistically significant.

When given psychometric data, clinicians and graduate students were more accurate than lay judges (e.g., undergraduates, secretaries) depending on the type of test data. Psychologists were not more accurate than lay judges when they were given results from projective tests, including results from the Rorschach Inkblot Method and Human Figure Drawings (Cressen, 1975; Gadol, 1969; Hiler & Nesvig, 1965; Levenberg, 1975; Schaeffer, 1964; Walker & Linden, 1967). Nor were clinical psychologists more accurate than lay judges when the task was to use screening instruments (e.g., the Bender–Gestalt test) to detect neurological impairment (Goldberg, 1959; Leli & Filskov, 1981, 1984; Nadler, Fink, Shontz, & Brink, 1959; Robiner, 1978). For example, in one of these studies (Goldberg, 1959), clinical psychologists were not more accurate than their own secretaries. Finally, when given MMPI protocols, psychologists and graduate students were more accurate than lay judges (Aronson & Akamatsu, 1981; Goldberg, 1968; Oskamp, 1962). For example, Aronson and Akamatsu (1981) compared the ability of graduate and undergraduate students to perform Q-sorts to describe the personality characteristics of patients with psychiatric conditions on the basis of MMPI protocols. Students' level of training differed in that graduate students had taken coursework in the MMPI and had some experience administering and/or interpreting the instrument, whereas undergraduates had only attended two lectures on the MMPI. Criterion ratings were based on family and patient interviews. Correlations between judges' ratings and criterion ratings were .44 and .24 for graduate and undergraduate students' ratings, respectively. Graduate student ratings were significantly more accurate than undergraduate ratings.

Clinicians Differing in Experience and Specialized Training

Mental health professionals who have received special training have been more accurate than other mental health professionals. For example, neuropsychologists are more accurate than clinical psychologists when the task is to detect neurological impairment (e.g., Goldstein, Deysach, & Kleinknecht, 1973); psychologists with specialized gerontological training are more likely to make essential age-related diagnoses and recommendations (Hillman, Stricker, & Zweig, 1997); psychologists with a background in forensic psychology are more likely than other psychologists to detect

lying (Ekman, O'Sullivan, & Frank, 1999); and psychiatrists do better than other physicians at prescribing antidepressant medicine (e.g., making sure a patient is on a therapeutic dose; Fairman, Drevets, Kreisman, & Teitelbaum, 1998). Similarly, in another study (Meyerson, Moss, Belville, & Smith, 1979), psychiatry residents did not change their decision making regarding hospitalization and medication due to a general increase in experience, but did so only when given specific clinical training.

Summary and Discussion

Clinicians who use pseudoscientific methods often do not discover that these methods are invalid because it is difficult to learn from clinical experience. Judgments and decisions made by experienced clinicians are generally no more accurate than those made by less experienced clinicians when all of the judges are given identical sets of information. Similarly, experienced clinicians are generally no more accurate than interns and advanced graduate students. On the other hand, clinicians and graduate students have outperformed lay judges, and in some instances mental health professionals with specialized training and interests outperform other mental health professionals. Obviously, longitudinal results that show that graduate students become more accurate after didactic training but not after training at a practicum site also suggest that training is of value but that it can be difficult to learn from clinical experience. To obtain a deeper understanding of why clinicians continue to use pseudoscientific methods, it is important to understand why it is difficult to learn from clinical experience.

IMPEDIMENTS TO LEARNING FROM EXPERIENCE

There are many reasons why it can be difficult for mental health professionals to learn from experience. Psychologists and other mental health workers are routinely confronted with ambiguous and complex decision-making tasks in which they must interpret and manipulate large amounts of data to formulate diagnostic impressions. Research has shown that in such situations professionals are susceptible to numerous cognitive and environmental influences that can result in poor judgments and a failure to learn from experience (Arkes, 1981; Brehmer, 1980; Dawes, 1994; Dawes, Faust, & Meehl, 1989; Einhorn, 1988; Garb, 1998, 2005; Tracey et al., 2014).

These cognitive factors include biases, heuristics, and memory processes, and environmental influences include the unavailability of adequate and appropriate feedback. These cognitive and environmental influences can exert a major negative impact on clinical care. Erroneous cognitive processes and poor feedback systems often result in the use of suboptimal hypothesis testing and decision-making strategies. When poor decision-making strategies are used, the likelihood that mental health workers will

make accurate judgments and effectively learn from experience is diminished.

Cognitive Processes and Errors in Judgment

Biases, heuristics, and memory processes can compromise the ability of clinicians to use optimal decision-making strategies. Biases are beliefs or preconceptions that adversely influence clinicians' interpretations of available data. Heuristics are simple rules or shortcuts that describe how clinicians make judgments, treatment decisions, or both. Reliance on heuristics can be efficient, but they are fallible and can lead clinicians to fail to learn from their experiences.

Biases include (but are not limited to) confirmatory bias, overconfidence, hindsight bias, race bias, gender bias, and bias to perceive psychopathology. Confirmatory bias occurs when clinicians knowingly or unknowingly review patient information in such a way that they seek and attend to information that can support but not counter their initial hypotheses. For example, in one study (Haverkamp, 1993), when psychology graduate students watched a videotape of an initial counseling session and listed the questions they would like to ask the client, their style of hypothesis testing was confirmatory 64% of the time (their questions could confirm but not refute their impressions of the client), neutral 21% of the time, and disconfirmatory only 15% of the time.

As revealed in another study, confirmatory information-seeking strategies occur early in the diagnostic decision-making process and can minimize the likelihood that clinicians will make accurate judgments. Gauron and Dickinson (1969) asked psychiatrists to describe their diagnostic impressions of patients on the basis of videotaped interviews. Many diagnoses were made within 30–60 seconds of viewing the interview and most never changed, even when the psychiatrists received disconfirming evidence. Presumably, because the clinicians did not actively seek out and attend to disconfirming evidence, they were less apt to change their initial opinions. When the psychiatrists received disconfirming evidence, they may have ignored it or, even more likely, they may have reinterpreted the contradictory evidence so that it became consistent with their initial beliefs.

Confirmatory bias can lead to overconfidence, and overconfidence can in turn lead to a greater reliance on confirmatory bias. Oskamp (1965) conducted a study in which the combined influence of overconfidence and confirmatory strategies was highlighted. Oskamp investigated the relation between incremental data gathering, validity of personality judgments, and confidence in the accuracy of judgments. Interestingly, clinicians' confidence in their personality judgments increased with the presentation of additional case history information. However, as confidence increased, accuracy remained at roughly the same level. Oskamp's data suggest that overconfidence led practitioners to emphasize information that confirmed

their initial hypotheses and to ignore or reinterpret information that did not support these hypotheses. The combination of confirmatory hypothesis-testing strategies and overconfidence may have prevented clinicians from attending to relevant patient data and contributed to diagnostic inaccuracy.

Overconfidence has been studied in two ways, by asking mental health professionals to rate their overall clinical skills compared to other mental health professionals and by having them make ratings for individual clients. When asked to compare their overall clinical skills to other mental health professionals, 25% viewed their skill level to be at the 90th percentile and none viewed themselves to be below average (Walfish, McAlister, O'Donnell, & Lambert, 2012). When making ratings for individual clients, psychologists have been overconfident in some studies (e.g., Oskamp, 1965) but underconfident in others (Wedding, 1983). In general, the greater the validity of assessment information, the stronger the relation between confidence and validity. Put another way, if the validity of judgments is not statistically significant, then one should not expect the relation between confidence and validity to be statistically significant.

Confirmatory bias and overconfidence can help to explain what goes wrong once a clinician has formed a hypothesis, but they do not help us understand how hypotheses are initially formulated. Consider the task of diagnosis. Clinicians may form hypotheses when they learn that a client meets at least some of the criteria for a diagnosis, but insidious influences are also sometimes at work. Clinicians can be influenced by a number of biases, such as race bias, social class bias, gender bias, age bias, labeling bias, and a tendency to overpsychopathologize clients (Garb, 2010). Thus, when interviewing an African American client, they may consider a diagnosis of schizophrenia before they consider a diagnosis of bipolar (manic–depressive) disorder. When clinicians engage in confirmatory hypothesis testing and become overconfident, they are unlikely to overcome the biases that led to their formulating a hypothesis.

Research on illusory correlations can also help us understand how hypotheses are initially formulated. This area of research also demonstrates that it can be difficult for clinicians to learn from clinical experience and that once clinicians formulate a hypothesis they tend to hold onto it despite disconfirming evidence. An illusory correlation occurs when a clinician believes there is a correlation between events that are not truly correlated, are only weakly correlated, or are correlated in the opposite manner from that assumed by the clinician.

In a pioneering study, Chapman and Chapman (1967) attempted to determine why clinicians continue to use the Draw-A-Person test in spite of a large body of research documenting no relation between picture characteristics (e.g., peculiar or emphasized eyes) and personality characteristics (e.g., suspiciousness) (see also Hunsley, Lee, Wood, & Taylor, Chapter 3, this volume). Chapman and Chapman conducted a two-part study. First, they asked clinicians to identify which features of drawings are associated

with particular symptoms and traits. Second, they presented human figure drawings to undergraduates. Undergraduates were to examine each drawing and then read a statement on the back that described a symptom or trait that was said to be possessed by the client who had drawn the picture. Undergraduates were unaware that the drawings and statements on the back of the drawings were randomly paired. Remarkably, undergraduates reported observing the same relations that had been reported by the clinicians, even though these relations were nonexistent in the stimulus materials. These results demonstrate that clinicians respond to verbal associations (e.g., between "eyes" and "watchfulness or suspiciousness"). Typically, they do not respond to the actual co-occurrence of particular picture features and psychiatric symptomatology (in most cases, picture features are unrelated to psychiatric symptomatology, even though some clinicians think that they are related; Groth-Marnat & Roberts, 1998; Kahill, 1984; Motta, Little, & Tobin, 1993; Thomas & Jolley, 1998).

Hindsight bias describes the mental processes that occur when individuals generate explanations for events that have occurred. For example, when clinicians learn that a client has committed suicide, they are likely to make high ratings if asked to rate the likelihood of the client committing suicide given the information they had about the client before the client committed suicide. Clinicians are generally unaware that the knowledge of an outcome influences the perceived likelihood of that outcome (Fischhoff, 1975). To put it another way, after an event has occurred, people are likely to believe that the event was bound to occur. This finding has been widely replicated across a range of judgment tasks (Hawkins & Hastie, 1990), such as the diagnosis of neurological impairment (Arkes, Faust, Guilmette, & Hart, 1988). Hindsight bias is important for understanding why clinicians have difficulty learning from clinical experience because it suggests that clinicians think in deterministic (not probabilistic) terms. In reality, all assessment information is fallible, and we frequently cannot make predictions with a high degree of certainty. To think otherwise is likely to lead clinicians to the erroneous belief that a particular combination of symptoms or behaviors is almost invariably associated with a particular outcome.

Additional errors in judgment can result from the use of heuristics. As already noted, heuristics are simple rules that describe how people make judgments (Kahneman, Slovic, & Tversky, 1982). They allow people to make judgments with a minimum of time and effort, and they frequently lead to accurate judgments. Nevertheless, in certain circumstances, they can also lead to mistakes in judgment.

The heuristic that is most relevant to understanding why clinicians can have difficulty learning from experience is the availability heuristic. The availability heuristic describes the role of memory in judgmental error. Clinicians may have trouble learning from experience because of the way clinicians remember or misremember information. Given that it is difficult, or even impossible, to retain all details of cases and clients, clinicians often

recall only selected information about each case. The recalled information, however, may not adequately describe the case or may be irrelevant to its central features.

According to the availability heuristic, a clinician's judgments are influenced by the ease with which the clinician remembers particular cases. The ease with which information can be remembered is related to several factors. For example, clinicians are more likely to remember a case that is striking, vivid, or unusual in some way. Similarly, clinicians often make judgments by forming causal theories that help them understand a client (Garb, 2005), and they are more likely to remember information that supports those theories (Kim & Ahn, 2002). Finally, clinicians are more likely to remember instances when a test indicator and symptom were present than cases when a test indicator was absent and a symptom was either present or absent (Arkes, 1981; Kayne & Alloy, 1988). To accurately determine how two events covary, one has to remember instances when the symptom or behavior did not occur as well as instances when they did occur. Of course, when clinicians cannot accurately determine how two events covary, an illusory correlation is said to be present.

In addition to the inherent fallibility of memory, Arkes and Harkness (1980) suggested that the act of making a diagnosis can influence how one remembers a client's symptoms. That is, a clinician's memory can be altered by the very act of making a diagnosis. A psychologist is likely to "recall" a patient having a symptom that is typical of the diagnosis, even though the client may not have that symptom. Similarly, the psychologist may forget that a client has a particular symptom because the symptom is atypical of the diagnosis rendered. Clinicians are unlikely to learn from experience when the details of cases are remembered incorrectly.

The Nature of Feedback in Clinical Practice

Mental health professionals typically do not receive feedback on the validity of their judgments and decisions. In most cases, to determine the accuracy of a judgment or decision, one would have to collect longitudinal or outcome data. This is accomplished in empirical studies, but most clinicians find it to be too expensive and time consuming to perform in clinical practice.

This state of affairs can be contrasted with most areas of medicine. Physicians frequently receive accurate feedback from laboratory results, radiology studies, and, in some cases, autopsies. These outcome measures are significantly more objective than those commonly used for mental health interventions (e.g., client reports of improved functioning). The paucity of appropriate outcome measures for psychology and psychiatry reflects both the difficulties of obtaining criterion information for mental health tasks and a lack of training in outcome evaluation techniques.

When mental health clinicians receive feedback by talking with a client,

the feedback is sometimes misleading. For example, a clinician may seek feedback from a client to determine if a test report is accurate or if a treatment has caused a reduction in distress. This method of obtaining feedback can be problematic for several reasons. First, clients may be reluctant to dispute their therapists' hypotheses. This may be due to passivity or suggestibility, fear of authority, or social desirability. Second, clients may be unaware of, or unable to describe, some of their own traits. Similarly, their reports of whether they have improved may be powerfully influenced by how they feel when they are asked. Third, psychological test reports often describe traits that are common to people in general, but do not describe traits that are specific to a client (e.g., "You are sensitive to other people's needs," "You have a great deal of unused potential," "You occasionally have difficulty making decisions"). Clients may accept the validity of the test report on face value.

This phenomenon has been labeled the P. T. Barnum effect, after the circus entrepreneur P. T. Barnum (Meehl, 1956), who once said that "I try to give a little something to everyone." Thus, a client may feel that a test report is accurate even though the report yields little individualized information. A client may be even more likely to believe that a test report is accurate if the report includes positive Barnum statements (this is termed the "Pollyanna principle"). For the client to say, "Yes, that describes me!" would be misleading to clinicians who are trying to understand the validity of their case conceptualizations, test reports, or both.

The Barnum effect was illustrated in a study by Logue, Sher, and Frensch (1992). They administered a personality inventory to 224 undergraduates, 112 of whom were adult children of alcoholics. Participants were informed that the purpose of the study was to validate a new personality questionnaire. Following completion of the questionnaire, participants were given a personality profile. They were told the profile was based on their responses to the test items, and they were asked to rate the accuracy of the profile description. Participants were randomly given one of two profiles, regardless of their test responses. The first was a general profile that was descriptive of many individuals (the Barnum profile), and the second was an "Adult Child of an Alcoholic" profile (based on commonly assumed characteristics of adult children of alcoholics). Examples of statements for the Barnum and Adult Child of an Alcoholic profiles are, respectively, "You are able to take criticism occasionally" and "In times of crisis you tend to take care of others."

Both profiles received widespread endorsement by all subjects. Of those who received the Barnum profile, 79% of the adult children of alcoholics and 70% of the remaining participants reported that the profile described them very well or better. Of those who received the Adult Child of an Alcoholic profile, 71% of the adult children of alcoholics and 63% of the remaining population reported that the profile described them very well

or better. Clearly, participants were unaware that test profiles had been randomly assigned to them and that the profiles were not really designed to describe their specific personality traits.

Clinicians also receive misleading feedback when they make incorrect interpretations but convince their clients that they are correct. For example, some clinicians tell their clients that they believe they were abused, even when their clients have no memory of having been abused. To help them "remember" having been abused, therapists may tell them that they were abused and repeatedly ask them to remember the events. Similarly, they may interpret their dreams, hypnotize them, or refer them to incest-survivor groups, all in an effort to help them "remember" an episode of abuse that may have never occurred. Too often, clients falsely remember having been abused (Loftus, 1993; Ofshe & Watters, 1994; see also Lynn, Krackow, Loftus, Locke, & Lilienfeld, Chapter 8, this volume).

SUMMARY AND DISCUSSION

Clinicians who use pseudoscientific assessment and treatment methods continue to use them in part because they have not learned from clinical experience that they do not work. Empirical studies may raise questions about the validity and utility of pseudoscientific methods, but clinical experience is less likely to do so. A great deal of research has been conducted on cognitive processes, the nature of feedback, and the reasons it is difficult to learn from experience. However, empirical studies have not focused on studying clinicians who use pseudoscientific methods. That is, studies have not focused on understanding why clinicians who use pseudoscientific methods have trouble learning from experience. Put another way, no study has looked at the cognitive processes, personality traits, and belief structures (cognitive schemas) of clinicians who use pseudoscientific methods. Nor has any study looked at social factors that may reinforce clinicians for using pseudoscientific methods.

One can wonder whether individual differences exist among clinicians in their attraction to pseudoscientific methods. For example, if a clinician is attracted to one pseudoscientific method, will the clinician also be attracted to others? Will there be some pseudoscientific methods that the clinician is not attracted to? If individual differences are present, then studies need to be conducted and theories need to be developed to help us understand these differences. Similarly, it would also be interesting to learn if some clinicians who use pseudoscientific methods graduate from training programs that reject pseudoscientific methods.

It is surprisingly difficult for clinicians to learn from clinical experience. This is not to say that clinical experience is never valuable. For example, experience may help clinicians structure judgment tasks (Brammer,

2002). That is, it may help clinicians decide what judgments and decisions need to be made. Similarly, more experienced clinicians may be better at knowing what information to collect in an interview. However, experience does *not* seem to be useful for helping clinicians evaluate the validity of an assessment instrument. Nor does experience seem to help clinicians make more valid judgments when the task is structured for them (e.g., when they are all given the same information).

Because it is difficult to learn from clinical experience, mental health professionals should not use an assessment instrument or treatment method solely because it seems to work in clinical practice. Instead, clinicians should become familiar with the research literature to learn if the assessment instrument or treatment method is supported by empirical research.

Mental health professionals also need to become familiar with the research literature on clinical judgment. By becoming familiar with the results of these studies, clinicians can avoid making judgments for tasks that are surprisingly difficult and for which they are unlikely to be accurate.

To bring about positive change, admissions policies may need to be changed for graduate school programs. Snyder (1995) addressed this issue:

> Are there forces attracting students to graduate training who are not predisposed toward scientific approaches to clinical psychology? . . . Unless changes occur in the type of students recruited to our field and in the work environment for graduates who become practicing clinicians, my sense is that all the efforts exerted in training programs, licensure, and continuing education will not accomplish any greater emphasis on the scientific principles underlying the helping process. (p. 423)

Thus, when selecting students for graduate school, one should select for scientific-mindedness as well as for intelligence and personality factors (e.g., warmth and empathy). Students should be selected who are likely to use an assessment instrument or a treatment intervention because it has been repeatedly supported by research, not because a charismatic figure has praised it or because in their experience they have found it to be helpful.

Additional changes will need to be made if empirically supported assessment instruments and treatment interventions are to be used more widely. For example, in addition to making changes to training programs, licensure exams should be critically appraised. Specifically, licensing exams should evaluate applicants' knowledge of empirically supported assessment instruments and treatment interventions. Only when such changes are made will clinicians become less likely to use pseudoscientific methods.

In conclusion, clinical lore suggests that psychologists "learn from doing" or learn from experience. However, empirical research indicates that it is difficult to learn from experience. To become more accurate, psychologists need to understand why it can be difficult to learn from experience, and they need to place greater emphasis on scientific findings.

GLOSSARY

Availability heuristic: The tendency for an individual's judgments to be influenced by the ease with which the individual remembers particular cases. This heuristic can lead to incorrect judgments. The ease with which information can be remembered is related to several factors (e.g., the strength of verbal associative connections between events may influence how well one "remembers" the co-occurrence of those events).

Biases: Beliefs or preconceptions that adversely influence clinicians' interpretations of available data. Also, cognitive processes that relate to the occurrence of cognitive biases (e.g., overconfidence, confirmatory bias).

Confirmatory bias: The tendency to seek, attend to, or remember information that can support but not counter a belief or preconception.

Gender bias: This bias occurs when, for a particular task, judgments are more valid for one gender than for another.

Heuristics: Simple rules or shortcuts by which individuals, including clinicians, make judgments, decisions, or both.

Hindsight bias: The tendency for knowledge of a particular outcome event to influence the perceived likelihood of that event, leading to incorrect explanations for the occurrence of the event.

Illusory correlation: The belief that events are correlated when in reality the events are not correlated, are only weakly correlated, or are correlated in the opposite direction from that assumed by the clinician.

Q-sort: A judgment task in which clinicians are instructed to describe a client by sorting items into categories ranging from least characteristic to most characteristic of the client. Items typically describe personality traits, psychiatric symptoms, or both.

Race bias: A bias that occurs when, for a particular task, judgments are more valid for one racial group than for another.

AUTHOR NOTE

The views expressed in this chapter are those of the authors and are not the official policy of the Department of Defense or the United States Air Force.

REFERENCES

Anthony, N. (1968). The use of facts and cues in clinical judgments from interviews. *Journal of Clinical Psychology, 24*, 37–39.

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49*, 323–330.

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73*, 305–307.

Arkes, H. R., & Harkness, A. R. (1980). Effect of making a diagnosis on subsequent recognition of symptoms. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 568–575.

Aronson, D. E., & Akamatsu, T. J. (1981). Validation of a Q-sort task to assess MMPI skills. *Journal of Clinical Psychology*, *37*, 831–836.

Brammer, R. (2002). Effects of experience and training on diagnostic accuracy. *Psychological Assessment, 14*, 110–113.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, *45*, 223–241.

Brenneis, C. B. (1971). Factors affecting diagnostic judgments of manifest dream content in schizophrenia. *Psychological Reports, 29*, 811–818.

Brenner, D., & Howard, K. I. (1976). Clinical judgment as a function of experience and information. *Journal of Clinical Psychology*, *32*, 721–728.

Chandler, M. J. (1970). Self-awareness and its relation to other parameters of the clinical inference process. *Journal of Consulting and Clinical Psychology*, *35*, 258–264.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, *72*, 193–204.

Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: The relationship between clinical experience and quality of health care. *Annals of Internal Medicine, 142*, 260–273.

Cressen, R. (1975). Artistic quality of drawings and judges' evaluations of the DAP. *Journal of Personality Assessment*, *39*, 132–137.

Danet, B. N. (1965). Prediction of mental illness in college students on the basis of "nonpsychiatric" MMPI profiles. *Journal of Consulting Psychology*, *29*, 577–580.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668–1674.

Ebling, R., & Levenson, R. W. (2003). Who are the marital experts? *Journal of Marriage and Family, 65*, 130–142.

Einhorn, H. J. (1988). Diagnosis and causality in clinical and statistical prediction. In D. C. Turk & P. Salovey (Eds.), *Reasoning, inference, and judgment in clinical psychology* (pp. 51–70). New York: Free Press.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10*, 263–266.

Fairman, K. A., Drevets, W. C., Kreisman, J. J., & Teitelbaum, F. (1998). Course of antidepressant treatment, drug type, and prescriber's specialty. *Psychiatric Services, 49*, 1180–1186.

Falvey, J. E., & Hebert, D. J. (1992). Psychometric study of the Clinical Treatment Planning Simulations (CTPS) for assessing clinical judgment. *Journal of Mental Health Counseling, 14*, 490–507.

Faust, D., Guilmette, T. J., Hart, K., Arkes, H. R., Fishburne, F. J., & Davey, L. (1988). Neuropsychologists' training, experience, and judgment accuracy. *Archives of Clinical Neuropsychology*, *3*, 145–163.

Fischhoff, B. (1975). Hindsight–foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288–299.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101*, 75–90.

Gadol, I. (1969). The incremental and predictive validity of the Rorschach test in personality assessments of normal, neurotic, and psychotic subjects. *Dissertation Abstracts*, *29*, 3482–B. (University Microfilms No. 69–4469)

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, *105*, 387–396.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89.

Garb, H. N. (2010). The social psychology of clinical judgment. In J. E. Maddux & J. P. Tangney (Eds.), *Social psychological foundations of clinical psychology* (pp. 297–311). New York: Guilford Press.

Garb, H. N., & Schramke, C. J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analyses. *Psychological Bulletin, 120*, 140–153.

Garner, A. M., & Smith, G. M. (1976). An experimental videotape technique for evaluating trainee approaches to clinical judging. *Journal of Consulting and Clinical Psychology*, *44*, 945–950.

Gaudette, M. D. (1992). Clinical decision making in neuropsychology: Bootstrapping the neuropsychologist utilizing Brunswik's lens model (doctoral dissertation, Indiana University of Pennsylvania, 1992). *Dissertation Abstracts International*, *53*, 2059B.

Gauron, E. F., & Dickinson, J. K. (1969). The influence of seeing the patient first on diagnostic decision making in psychiatry. *American Journal of Psychiatry, 126*, 199–205.

Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender–Gestalt test. *Journal of Consulting Psychology, 23*, 25–33.

Goldberg, L. R. (1965). Diagnosticians versus diagnostic signs: The diagnosis of psychosis versus neurosis from the MMPI. *Psychological Monographs*, *79*(9, Whole No. 602).

Goldberg, L. R. (1968). Simple models or simple processes?: Some research on clinical judgments. *American Psychologist*, *23*, 483–496.

Goldstein, S. G., Deysach, R. E., & Kleinknecht, R. A. (1973). Effect of experience and amount of information on identification of cerebral impairment. *Journal of Consulting and Clinical Psychology*, *41*, 30–34.

Graham, J. R. (1967). A Q-sort study of the accuracy of clinical descriptions based on the MMPI. *Journal of Psychiatric Research*, *5*, 297–305.

Graham, J. R. (1971). Feedback and accuracy of clinical judgments from the MMPI. *Journal of Consulting and Clinical Psychology*, *36*, 286–291.

Grebstein, L. (1963). Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. *Journal of Consulting Psychology, 37*, 127–132.

Griswold, P. M., & Dana, R. H. (1970). Feedback and experience effects on psychological reports and predictions of behavior. *Journal of Clinical Psychology*, *26*, 439–442.

Groth-Marnat, G., & Roberts, L. (1998). Human Figure Drawings and House-Tree-Person drawings as indicators of self-esteem: A quantitative approach. *Journal of Clinical Psychology, 54*, 219–222.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review*, *71*, 438–456.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*, 155–163.

Haverkamp, B. E. (1993). Confirmatory bias in hypothesis testing for client-identified and counselor self-generated hypotheses. *Journal of Counseling Psychology, 40*, 303–315.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311–327.

Heaton, R. K., Smith, H. H., Jr., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*, 892–900.

Hermann, R. C., Ettner, S. L., Dorwart, R. A., Langman-Dorwart, N., & Kleinman, S. (1999). Diagnoses of patients treated with ECT: A comparison of evidence-based standards with reported use. *Psychiatric Services, 50*, 1059–1065.

Hiler, E. W., & Nesvig, D. (1965). An evaluation of criteria used by clinicians to infer pathology from figure drawings. *Journal of Consulting Psychology*, *29*, 520–529.

Hillman, J. L., Stricker, G., & Zweig, R. A. (1997). Clinical psychologists' judgments of older adult patients with character pathology: Implications for practice. *Professional Psychology: Research and Practice, 28*, 179–183.

Holmes, C. B., & Howard, M. E. (1980). Recognition of suicide lethality factors by physicians, mental health professionals, ministers, and college students. *Journal of Consulting and Clinical Psychology*, *48*, 383–387.

Horowitz, M. J. (1962). A study of clinicians' judgments from projective test protocols. *Journal of Consulting Psychology*, *26*, 251–256.

Jackson, R. L., Rogers, R., & Shuman, D. W. (2004). The adequacy and accuracy of sexually violent predator evaluations: Contextualized risk assessment in clinical practice. *International Journal of Forensic Mental Health, 3*, 115–129.

Johnston, R., & McNeal, B. F. (1967). Statistical versus clinical prediction: Length of neuropsychiatric hospital stay. *Journal of Abnormal Psychology*, *72*, 335–340.

Kahill, S. (1984). Human figure drawing in adults: An update of the empirical evidence, 1967–1982. *Canadian Psychology, 25*, 269–292.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kayne, N. T., & Alloy, L. B. (1988). Clinician and patient as aberrant actuaries: Expectation-based distortions in assessment of covariation. In L. Y. Abramson (Ed.), *Social cognition and clinical psychology: A synthesis* (pp. 295–365). New York: Guilford Press.

Kendell, R. E. (1973). Psychiatric diagnoses: A study of how they are made. *British Journal of Psychiatry*, *122*, 437–445.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131*, 451–476.

Lambert, M. J. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire)*. Salt Lake City, UT: OQ Measures.

Lambert, L. E., & Wertheimer, M. (1988). Is diagnostic ability related to relevant training and experience? *Professional Psychology: Research and Practice, 19*, 50–52.

Leli, D. A., & Filskov, S. B. (1981). Clinical-actuarial detection and description of brain impairment with the W-B Form I. *Journal of Clinical Psychology*, *37*, 623–629.

Leli, D. A., & Filskov, S. B. (1984). Clinical detection of intellectual deterioration associated with brain damage. *Journal of Clinical Psychology*, *40*, 1435–1441.

Levenberg, S. B. (1975). Professional training, psychodiagnostic skill, and Kinetic Family Drawings. *Journal of Personality Assessment*, *39*, 389–393.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48*, 518–537.

Logue, M. B., Sher, K. J., & Frensch, P. A. (1992). Purported characteristics of adult children of alcoholics: A possible "Barnum Effect." *Professional Psychology: Research and Practice, 23*, 226–232.

Meehl, P. E. (1956). Wanted—a good cookbook. *American Psychologist, 11*, 263–272.

Meehl, P. E. (1997). Credentialed persons, credentialed knowledge. *Clinical Psychology: Science and Practice, 4*, 91–98.

Meyerson, A. T., Moss, J. Z., Belville, R., & Smith, H. (1979). Influence of experience on major clinical decisions: Training implications. *Archives of General Psychiatry, 36*, 423–427.

Morgan, R. D., Olson, K. R., Krueger, R. M., Schellenberg, R. P., & Jackson, T. T. (2000). Do the DSM decision trees improve diagnostic ability? *Journal of Clinical Psychology, 56*, 73–88.

Motta, R. W., Little, S. G., & Tobin, M. I. (1993). The use and abuse of human figure drawings. *School Psychology Quarterly, 8*, 162–169.

Muller, M. J., & Davids, E. (1999). Relationship of psychiatric experience and interrater reliability in assessment of negative symptoms. *Journal of Nervous and Mental Diseases, 187*, 316–318.

Nadler, E. B., Fink, S. L., Shontz, F. C., & Brink, R. W. (1959). Objective scoring vs. clinical evaluation of the Bender-Gestalt. *Journal of Clinical Psychology, 15*, 39–41.

Ofshe, R., & Watters, E. (1994). *Making monsters: False memories, psychotherapy, and sexual hysteria*. New York: Scribner's.

Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs*, *76*(28, Whole No. 547).

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, *29*, 261–265.

Robiner, W. N. (1978). *An analysis of some of the variables influencing clinical use of the Bender-Gestalt*. Unpublished manuscript.

Schaeffer, R. W. (1964). Clinical psychologists' ability to use the Draw-A-Person Test as an indicator of personality adjustment. *Journal of Consulting Psychology*, *28*, 383.

Schinka, J. A., & Sines, J. O. (1974). Correlates of accuracy in personality assessment. *Journal of Clinical Psychology*, *30*, 374–377.

Snyder, C. R. (1995). Clinical psychology building inspection: Slipping off its

science-based foundation [Review of the book *House of cards: Psychology and psychotherapy built on myth*]. *Contemporary Psychology, 40*, 422–424.

Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2009). The Meta-Analysis of Clinical Judgment project: Effects of experience on judgment accuracy. *Counseling Psychologist, 37*, 350–399.

Stelmachers, Z. T., & McHugh, R. B. (1964). Contribution of stereotyped and individualized information to predictive accuracy. *Journal of Consulting Psychology*, *28*, 234–242.

Stricker, G. (1967). Actuarial, naive clinical, and sophisticated clinical prediction of pathology from figure drawings. *Journal of Consulting Psychology*, *31*, 492–494.

Thomas, G. V., & Jolley, R. P. (1998). Drawing conclusions: A re-examination of empirical and conceptual bases for psychological evaluation of children from their drawings. *British Journal of Clinical Psychology, 37*, 127–139.

Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69*, 218–229.

Turner, D. R. (1966). Predictive efficiency as a function of amount of information and level of professional experience. *Journal of Projective Techniques and Personality Assessment*, *30*, 4–11.

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110*, 1–6.

Walker, C. D., & Linden, J. D. (1967). Varying degrees of psychological sophistication in the interpretation of sentence completion data. *Journal of Clinical Psychology, 23*, 229–231.

Walker, E., & Lewine, R. J. (1990). Prediction of adult-onset schizophrenia from childhood home movies of the patients. *American Journal of Psychiatry, 147*, 1052–1056.

Walters, G. D., White, T. W., & Greene, R. L. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *Journal of Consulting and Clinical Psychology*, *56*, 111–117.

Wanderer, Z. W. (1969). Validity of clinical judgments based on human figure drawings. *Journal of Consulting and Clinical Psychology*, *33*, 143–150.

Watson, C. G. (1967). Relationship of distortion to DAP diagnostic accuracy among psychologists at three levels of sophistication. *Journal of Consulting Psychology*, *31*, 142–146.

Waxer, P. (1976). Nonverbal cues for depth of depression: Set versus no set. *Journal of Consulting and Clinical Psychology*, *44*, 493.

Wedding, D. (1983). Clinical and statistical prediction in neuropsychology. *Clinical Neuropsychology*, *5*, 49–55.

Weiss, J. H. (1963). The effect of professional training and amount and accuracy of information on behavioral prediction. *Journal of Consulting Psychology*, *27*, 257–262.

Whipple, J. L., & Lambert, M. J. (2011). Outcome measures for practice. *Annual Review of Clinical Psychology, 7*, 87–111.

Whitehead, W. C. (1985). Clinical decision making on the basis of Rorschach, MMPI, and automated MMPI report data (Doctoral dissertation, University

of Texas at Southwestern Medical Center at Dallas, 1985). *Dissertation Abstracts International, 46–08B*, 2828.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Witteman, C. L. M., & van den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment, 23*, 56–61.

CHAPTER THREE

Controversial and Questionable Assessment Techniques

John Hunsley, Catherine M. Lee,
James M. Wood, and Whitney Taylor

The past decade has seen many important developments in the field of clinical assessment. These include (1) statistical approaches for exploring consistency and variability in reliability estimates (e.g., Vacha-Haase, Henson, & Caruso, 2002), (2) theoretical and methodological advances in conceptualizing construct validity (e.g., Strauss & Smith, 2009), (3) a renewed focus on the utility of assessment data in the clinical enterprise (e.g., McFall, 2005), (4) a compelling, empirically based rationale for routinely monitoring the impact of clinical interventions (e.g., Lambert & Shimokawa, 2011), and (5) initial attempts to delineate the nature and implications of an evidence-based approach to assessment (e.g., Hunsley & Mash, 2007; see also Garb & Boyle, Chapter 2, this volume). Despite this progress, there is widespread use of clinical assessment practices and instruments that lack a strong scientific foundation. In this chapter, we first provide introductory comments on key scientific elements of clinical assessment, and then we examine a subset of commonly used instruments whose use is not justified by scientific evidence.

At the heart of a scientific approach to psychological assessment lie the principles of falsifiability and methodological skepticism (e.g., Alcock, 1991; Popper, 1959). These principles require that claims about the scientific merits or validity of a hypothesis, measure, or theory be framed in a way that they can be (1) subjected to empirical investigation (i.e., by data-based investigation, rather than by reliance on appeals to anecdotal evidence or to special knowledge or authority), (2) refuted or disconfirmed by

empirical investigation, and (3) independently investigated by both proponents and critics of the claims. Unfortunately, as we show later in the chapter, clinical use of a number of assessment instruments has far exceeded their scientific bases. We do not accept the frequent claim made by proponents of such instruments that the research base has not yet "caught up" with what clinicians know to be an instrument of unquestionable clinical value of an instrument, or that research evidence is somehow irrelevant to the enterprise of clinical assessment. Professional standards in psychology, across numerous jurisdictions and countries, all require that practice be based on the science of psychology. In the scientific enterprise, the burden of proof to demonstrate the validity of a tool rests with those who propose its use (Lett, 1990).

PSYCHOLOGICAL TESTS

Psychological assessment is not synonymous with psychological testing, but psychological tests commonly comprise a major part of the assessment process. By definition, a psychological test is the measurement of a sample of behavior obtained under standardized conditions and that has established rules for scoring or interpreting this sample (Anastasi & Urbina, 1997). There are thousands of psychological tests, which vary enormously in their scientific foundations and psychometric adequacy. In this chapter we examine several problematic psychological tests commonly used in clinical assessment.

Standards for psychological tests and for their appropriate professional use are well developed and widely known (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). These standards set out the criteria for the development and use of psychological tests. They are designed to ensure that test developers and test users meet consensually defined expectations, held by the profession and the public, with regard to the scientifically supported use of tests. Proponents of questionable and controversial tests frequently claim the legitimacy that is associated with scientifically sound measures. However, they also sometimes deny that the test should be subjected to the high standards expected of a psychological test because of its obvious clinical value, or that it is not "really" a test, but merely a method for collecting information.

TEST CONSTRUCTION AND PSYCHOMETRIC PRINCIPLES

We next focus on elements that are required for a test to be both psychometrically sound and clinically useful. These elements, which hold for all types of psychological tests, are standardization (of stimuli, administration,

and scoring), reliability, validity, and norm. Greater details on these and other psychometric elements can be found in Haynes, Smith, and Hunsley (2011) and Wood, Garb, and Nezworski (2006).

Standardization is the first step in ensuring that obtained results can be replicated by another assessor. Unless there is standardization, results may be specific to the unique aspects of the testing situation. Standardization minimizes the influence of unique aspects of the testing situation and the assessor. To this end, test developers typically provide detailed instructions regarding the nature of the stimuli, administrative procedures, time limits (if relevant), and the types of verbal probes and responses to the examinee's questions that are permitted. Instructions are provided for the scoring of the test. In some cases, only simple addition of responses is required to obtain a test result; many tests have complex scoring rules that are mastered through extensive training.

Reliability is the next criterion addressed in the development of a scientifically sound test. Reliability refers to consistency—whether (1) all aspects of the test contribute in a meaningful way to the data obtained (internal consistency), (2) similar results would be obtained if the test was conducted and/or scored by another evaluator (interrater reliability), and (3) similar results would be obtained if the person was retested after the initial test (retest reliability or test stability). Standardization of stimuli, administration, and scoring are necessary, but not sufficient, to establish reliability. Reliable results are crucial for generalizing the results beyond the immediate context of the assessment. Thorough and complete test standardization does not guarantee reliability. For example, the test may consist of too many components that are influenced by ephemeral characteristics of the examinee or by contextual characteristics of the testing, including demand characteristics associated with the purpose of the testing or the behavior of the examiner. Alternatively, scoring criteria may be too complicated or insufficiently detailed to ensure reliable scoring by different assessors. Reliability values are always conditional, based on the characteristics of the assessment activity, the context of the assessment, and the nature of the sample of individuals who completed the test (Haynes et al., 2011).

Validity addresses the issue of whether the test measures what it purports to measure. A standardized and reliable test does not necessarily yield valid data. Validity refers to whether the test (1) samples the type of behavior that is relevant to the purpose of the test (content validity), (2) provides data consistent with theoretical postulates associated with the phenomenon assessed (concurrent and predictive validity), and (3) provides a measure of the phenomenon that is minimally contaminated by other psychological phenomena (discriminant validity). In applied contexts, it is also essential to consider incremental validity—the extent to which data from a test add to our knowledge over and above the information gleaned from other data (Hunsley & Meyer, 2003). Although it is common to talk about a test as valid or invalid, validity, like reliability, is conditional. Many psychological

tests consist of subscales designed to measure discrete aspects of a more global construct. In such situations, it is erroneous to talk about the validity of the test because the validity of each subscale must be established. Moreover, it is erroneous to refer to global validity of a test or subscale because validity can only be established within certain parameters, such that a test may be valid for specific purposes within specific groups of people (e.g., specific ages or genders). Finally, if a test is used for multiple purposes, its validity for each purpose must be empirically established. For example, knowing that a self-report test of psychological distress is a valid indicator of diagnostic status does not automatically support its use for such forensic purposes as determining competency or child custody arrangements.

For a test to be clinically useful, it must meet the criteria of standardization, reliability, and validity. However, to meaningfully interpret the results obtained from a single individual, it is essential to have *norms*, *specific criterion-related cutoff scores*, or both (American Educational Research Association et al., 1999). The meaning of the test results can only be considered in light of such reference points. Knowing that a person scored low or high on a test (i.e., relative to the range of possible scores) provides no meaningful information. Comparisons must be made with either criteria that have been set for a test (e.g., a certain degree of accuracy as demonstrated in the test is necessary for the satisfactory performance of a job) or with population norms. It is a challenging task to select the target population(s) and establish norms. For example, are the norms to be used for comparing a specific score with those that might be obtained within the general population or within specific subgroups of this population (e.g., gender-specific norms), or are the norms to be used for establishing the likelihood of membership in specific categories (e.g., nondistressed versus psychologically disordered groups)? As with validity considerations, it may be necessary to develop multiple norms for a test, based on the group being assessed and the purpose of the assessment.

DISTINGUISHING BETWEEN VALID AND INVALID USE OF ASSESSMENT TECHNIQUES: OR, WHEN IS A TEST NOT A TEST?

The appropriate use of psychometrically sound tests requires that guidelines for administration and scoring are followed and that relevant reliability values, validity indices, and group norms are used to interpret the obtained data. The assessor must select a test that is valid for the assessment purpose and for the person being assessed.

Widely accepted definitions of psychological tests are designed to enable psychologists to determine whether an assessment technique is a test. To define an information-gathering activity as a psychological test, two conditions must be met. First, a sample of behavior is collected to generate statements about a person, a person's experiences, or a person's psychological

functioning. Second, a claim is made or implied that the accuracy or validity of these statements stem from the way in which the sample was collected (i.e., the nature of the stimuli, technique, or process that gave rise to the sample of behavior), not just from the expertise, authority, or qualifications of the assessor. When both conditions are met, we consider that the process used to collect and interpret the behavioral sample is presented as a psychological test. This is consistent with the standards for psychological tests, which emphasize that an assessment method that relies on or uses the concepts and techniques of psychological testing is a test (American Educational Research Association et al., 1999). Despite the apparent simplicity of this definition, proponents of questionable techniques sometimes employ the term "psychological test" loosely, arguing simultaneously that scientific and professional standards, expectations, and responsibilities are inapplicable to their techniques, while also claiming a valid approach that is supported by scientific evidence. Thus, the issue of whether a specific form of data collection constitutes a test is far from an ivory tower concern for semantic hair-splitting.

For example, the Rorschach Inkblot Test has been used by clinical psychologists for many decades. Some Rorschach proponents have argued that the Rorschach is not a psychological test, but that it is instead a method of interviewing that generates data relevant to the practice of clinical assessment (e.g., Weiner, 1994). A slight variation on this approach is to treat the Rorschach as a test (by using a recognized scoring system), but then "enriching" the test results with personalized, interpretive speculations stemming from selected aspects of the examinee's test responses (e.g., Acklin, 1995). Such positions allow the assessor to claim that there is scientific evidence supporting the use of the Rorschach, while freeing the assessor to use the data in a manner unconstrained by issues of administration, norms, reliability, or validity.

CONTROVERSIAL AND QUESTIONABLE ASSESSMENT TECHNIQUES: SOME EXAMPLES

Next, we present a number of problematic assessment techniques that, according to surveys of assessment practices, are used by large numbers of clinical psychologists. Although we could consider many other examples of unscientific psychological assessment techniques, we have chosen to focus on five questionable assessment techniques that continue to be routinely used by psychologists: the Rorschach Inkblot Test, the Thematic Apperception Test, projective drawings, anatomical dolls, and the Myers–Briggs Type Indicator. For a fascinating look at some of the history behind the creation and development of some of these instruments, see Paul (2004).

Psychologists in the domain of personality assessment have long distinguished between projective tests and self-report inventories (cf. Vitacco,

Lilienfeld, Erickson, & Wood, 2012). Projective tests such as the Rorschach or TAT generally present the person being tested with an ambiguous stimulus (such as an inkblot or a picture without a caption) and ask for an open-ended response to the stimulus (e.g., "What might this be?" or "What do you think is happening in this picture?"). In contrast, self-report inventories (sometimes referred to as objective tests) generally present the examinee with a statement (e.g., "I often feel like crying") and ask the person to choose among two or more responses to indicate the extent to which the statement accurately reflects the person's experience.

Among problematic techniques discussed in this chapter, the majority fit the definition of projective techniques. As we will demonstrate, problems of standardization are rife among projective techniques. Questionable techniques are not, however, limited to projectives. To illustrate this point, we also review a self-report personality measure, the MBTI. Although standardization is not an issue with this test, concerns about its reliability and validity highlight the need for clinicians to select tests that have firm scientific support.

Rorschach Inkblot Test

First published in Switzerland more than 90 years ago (Rorschach, 1921/1964), the Rorschach Inkblot Test is one of the oldest and most controversial tests in psychology. An examinee is shown 10 symmetrical inkblots—5 in black and white, 5 in color—and asked to describe what they might represent. According to proponents of the test, an examinee's responses to these ambiguous shapes provide important evidence regarding his or her psychological functioning.

From the 1930s through the 1960s, leading Rorschach proponents hotly disagreed among themselves about how the test should be administered, scored, and interpreted (Exner, 1969). At least four competing approaches or "systems" for the Rorschach were widely used, and many clinicians borrowed elements from all four to personalize the way the test was scored and interpreted (Exner & Exner, 1972). The test entered a new era in the 1970s, when John Exner introduced his Comprehensive System for the Rorschach (CS), which was said to incorporate the best parts from all the other systems (Exner, 1974). Regularly updated until Exner's death in 2006 (Exner, 2003; Exner & Erdberg, 2005), the CS became the prevailing system for teaching and researching the test (Hilsenroth & Handler, 1995; Shontz & Green, 1992). Today it remains the dominant approach to the Rorschach, although a new system, the Rorschach Performance Assessment System (R-PAS), was recently introduced (Meyer, Viglione, Mihura, Erard, & Erdberg, 2011).

In contrast to many Rorschach proponents (e.g., Weiner, 1994), Exner always insisted that the Rorschach is a psychological test and, as such, must meet the standards expected of a test (e.g., Exner, 1997). The developers

of the new R-PAS have taken a similar position (Meyer et al., 2011). As the most commonly researched and used projective measure (Camara, Nathan, & Puente, 2000; Butcher & Rouse, 1996; Watkins, Campbell, Nieberding, & Hallmark, 1995), the Rorschach has been the focus of a great deal of scientific attention over the past several decades. However, even Rorschach proponents have accepted the fact that much of the early research was poorly done or irrelevant to current uses of the test (Exner, 1986). Thus, we will focus primarily on research regarding the CS that has been published since the mid-1970s. Because only a handful of peer-reviewed studies have been published on the newly created R-PAS, it will not be further discussed here.

Standardization

The Rorschach is a time-consuming measure to administer, score, and interpret, requiring an average of approximately 2.5 hours of clinician time (Ball, Archer, & Imhof, 1994; Camara et al., 2000). The CS offers detailed rules for administration and scoring, with extensive tables to aid in the scoring and interpretation of the test results. Assessors are provided directions on the seating arrangement to be used, the sequence of card administration, and the instructions to be given to examinees.

Surveys regarding the use of psychological tests in clinical and legal settings indicate that the majority of clinical psychologists and approximately one-third of forensic psychologists use the Rorschach at least occasionally (Camara et al., 2000; Archer, Buffington-Vollum, Stredny, & Handel, 2006). Given this widespread usage, it is surprising that little attention has been given to studying examiners' fidelity to the CS administration rules. Exner (1986, 1993) claimed that learning to administer the Rorschach correctly was not difficult. To demonstrate his point, he recounted how dentists, social workers, and secretaries had all successfully learned to administer the test while working on his research projects. However, other Rorschach proponents (Viglione & Meyer, 2008, p. 38; see also Weiner, 2001) have expressed concern that ambiguity in the CS administration rules has sometimes led to serious errors, distorted test scores, and potentially misleading research results reported in influential scientific journals.

These concerns about CS administration errors received important scientific support from a study by Lis, Parolin, Calvo, Zennaro, and Meyer (2007). The Rorschach was administered to a sample of 212 adults by a research team of graduate students who had received standard training in the CS and were supervised by psychologists highly knowledgeable about CS procedures. The team members then received additional intensive training from an internationally recognized Rorschach expert, after which they administered the Rorschach to a second sample of 101 adults.

The study's central finding was that the two samples of adults gave dissimilar types of responses to the inkblots and received substantially

different CS scores, even though the samples were tested by the same researchers, were selected according to the same criteria, and were similar in respect to cultural and sociocultural background. Lis and her colleagues (2007) concluded that the Rorschach responses from the two samples were dissimilar because members of the research team had inadvertently made serious errors when administering the Rorschach to the first sample.

The Lis et al. (2007) study highlighted an important weakness of Rorschach administration: Errors are easy to make, even if the examiner is careful, has received graduate training in the CS, and is well supervised. Furthermore, these errors are likely to have a substantial impact on the examinee's Rorschach responses and test scores while remaining undetectable to the examiner, his or her supervisor, and anyone who reviews the examiner's work. Because the Rorschach is susceptible to these invisible administration errors, reliance on the test can be especially risky in clinical and forensic evaluations in which uncorrected mistakes can have serious negative consequences.

Even if CS administrative requirements are scrupulously followed, there is extensive evidence that relatively innocuous contextual factors in Rorschach administration, such as the layout of the testing room and the appearance of the assessor, affect examinees' responses (Masling, 1992). Thus, the bottom line with regard to the routine clinical and forensic use of the Rorschach is that there is no way to establish with a high degree of certainty that a Rorschach has been administered, scored, and interpreted according to CS standards or that the test data are free from contamination by examiner error or other extraneous influences.

Norms

Over the course of more than three decades, Exner published numerous norms for children and adults (e.g., Exner, 1993, 2003, 2007) that came to be regarded as a cornerstone of the scientific basis of his Rorschach system. However, between 1999 and the present, overwhelming evidence from more than 50 studies has accumulated showing that the CS norms for many important Rorschach variables are seriously in error (Hamel, Shaffer, & Erdberg, 2000; Meyer, Erdberg, & Shaffer, 2007; Shaffer, Erdberg, & Haroian, 1999; Wood, Nezworski, Garb, & Lilienfeld, 2001; see also Meyer, 1991). It is now widely recognized that Exner's norms have a strong tendency to "overpathologize." That is, psychologically healthy children and adults are likely to be misidentified by the Rorschach as psychologically disturbed. More than 50% of normal individuals who are evaluated with the CS norms will be mislabeled by the test as narcissistic, depressed, or suffering from near-psychotic cognitive distortions (Hamel et al., 2000; Meyer et al., 2007; Meyer & Viglione, 2008; Wood, Nezworski, Lilienfeld, & Garb, 2003, 2009).

Although Exner (2001b, p. 172) publicly acknowledged that a serious

computational mistake had been made when his norms were being compiled, he steadfastly denied that the norms were erroneous (Exner, 2001a). Toward the end of his life, however, as the norms came under ever-increasing attack, he began to collect new data. An updated set of CS norms was published after his death (Exner, 2007) but failed to resolve the controversy. Exner's posthumously published norms, like his prior ones, were inconsistent with findings from other researchers and tended to overpathologize normal children and adults (Meyer & Viglione, 2008).

The CS norms are so gravely flawed that their continuing use in clinical and forensic settings poses a serious risk to the well-being of patients and other examinees. Some Rorschach proponents have attempted to deal with the problem by creating a new set of "International Norms" for the Rorschach based on CS results from the United States and dozens of other countries (see Meyer et al., 2007; Meyer & Viglione, 2008). Although those involved in the project recommended the adult norms for clinical use, they explicitly cautioned against using the youth norms and called for additional work to develop accurate norms for children and adolescents. However, these International Norms for adults are also problematic because they are based on a mixture of individuals from numerous nations who were administered the test in many different languages. The sampling strategies used to recruit research participants varied enormously from country to country and were not developed with the goal of obtaining nationally representative samples within each country. Thus the International Norms do not accurately represent the typical Rorschach performance for any particular country, cultural group, or language. To use these norms in the evaluation of any particular individual—for example, an American who has taken the Rorschach in English—is likely to involve inappropriate and potentially misleading comparisons of test performance across nationalities, cultures, and languages.

Reliability

Because there is a longstanding debate among Rorschach scholars regarding the relevance of internal consistency in assessing the reliability of the Rorschach (Reznikoff, Aronow, & Rauchway, 1982; Meyer & Viglione, 2008), evaluations of reliability have usually focused on interrater and test–retest reliability. We will first discuss interrater reliability before turning to test–retest reliability.

For many years Exner claimed that the interrater reliability of all CS scores was .85 or higher (Exner, 1978, p. 14; 1986, p. 23). However, independent studies eventually revealed this claim to be a substantial overstatement. For example, the combined results from well-done studies by Acklin, McDowell, Verschell, and Chan (2000), McGrath and his colleagues (2005), and Sultan and his colleagues (Sultan, Andronikof, Réveillère, & Lemmel, 2006; Sultan & Meyer, 2009) show that only about one-half of

the key CS scores have an interrater reliability of .85 or higher as measured by the intraclass correlation coefficient (ICC). Although other studies (Meyer et al., 2002; Viglione & Taylor, 2003) have reported somewhat higher figures, their methodology and statistical analyses are problematic (Wood et al., 2003, pp. 231–234, 366–367).

Is the interrater reliability of CS scores acceptable? This question has several answers because psychologists disagree about the standards that should be applied to clinical and forensic measures. Rorschach proponents (e.g., McGrath et al., 2005) have typically expressed a preference for lenient standards of interrater reliability, arguing that an ICC of .60 or higher is adequate. In support of their position, they point out that a minimum reliability of .60 is generally considered acceptable for psychiatric diagnoses and clinical judgments (e.g., Fleiss, 1981).

In contrast, some assessment experts reject a .60 cutoff as too lax (Ayearst & Bagby, 2010; Garb, Lilienfeld, & Fowler, 2008). Although these experts agree that a reliability of .60 is adequate for diagnoses and other clinical judgments, they argue that such judgments cannot meet even this minimal standard of reliability unless they are based on high-quality information: "If a test cannot be scored reliably, then the judgments made based on those test scores will necessarily have poor reliability" (Garb et al., 2008, p. 104). Accordingly, to increase the chances of reliable test interpretations derived from individual scores, these experts recommend that the interrater reliability of clinical tests be at least .90.

Other assessment experts have offered recommendations that lie between the two positions just described. Nunnally and Bernstein (1994) recommended that the minimum reliability of tests be .80 or higher (although recommending a value of .90 or higher as appropriate for tests used for clinical assessment purposes), whereas Hunsley and Mash (2008) have suggested that an interrater reliability of .70 or higher may be adequate.

Given the diversity of opinions, the issue of Rorschach interrater reliability is probably best addressed as a series of questions. The answers to these questions were calculated by averaging numbers from several high-quality studies on CS interrater reliability (Acklin et al., 2000; McGrath et al., 2005; Sultan et al., 2006; Sultan & Meyer, 2009).

First, how many Rorschach scores meet the stringent standard of the .90 level of interrater reliability recommended by some experts (Ayearst & Bagby, 2010)? According to the studies we have cited, only 30% of the key CS scores (19 out of 64) meet this high standard. As a point of comparison, all of the Wechsler IQ subtests meet this standard (Psychological Corporation, 1997). There is no question that the interrater reliability of Rorschach scores is generally much lower than that of Wechsler intelligence test scores.

Second, how many Rorschach scores meet the somewhat more relaxed standards proposed by Nunnally and Bernstein (1994) and Hunsley and Mash (2008)? According to the studies on this topic, about 64% of key CS

scores (41 out of 64) have interrater reliability of .80 or higher, whereas about 89% (57 out of 64) have reliability of .70 or higher. Third and finally, how many CS scores meet the lenient standards recommended by Rorschach proponents (e.g., McGrath et al., 2005)? All but five CS scores meet this minimal standard. As can be seen, the relatively low standards advocated by Rorschach proponents allow them to claim that virtually all CS scores have adequate reliability.

We turn next to the test–retest reliability of CS Rorschach scores. Until recently, this important psychometric topic, like much else about the CS, was clouded by conflicting and ambiguous information. During the past 15 years, Rorschach proponents repeatedly claimed that the average test–retest reliability of CS scores was in the high .70s or low.80s (Gronnerod, 2003; Meyer, 1997; Meyer & Archer, 2001; Viglione & Hilsenroth, 2001), which would be comparable to the temporal stability of scores on the MMPI-2 and well-accepted intelligence tests. However, these optimistic claims were based on fragmentary data, as critics pointed out (Garb, Wood, Nezworski, Grove, & Stejskal, 2001). A complete and methodologically adequate study of CS test stability was not published until 2006, when Sultan and his colleagues (2006; see also Sultan & Meyer, 2009) reported findings for a sample of 75 French nonpatient adults.

To these researchers' surprise, they found that the average test–retest reliability of CS scores over a three-month interval was in the low .50s, which was dramatically lower than Rorschach proponents had been claiming and far below the stability figures for the MMPI-2 and intelligence tests. Many CS scores with low stability had long been held forth as measures of stable personality traits and long-term behavioral dispositions. However, in the Sultan et al. (2006) study, only 9 (19%) of 47 core Rorschach variables had test–retest reliability of .70 or higher, which would be regarded as a minimal standard for scores that purport to measure stable personality characteristics (Hunsley & Mash, 2008). These findings raise questions regarding claims by Rorschach proponents that the test can detect enduring trait-like dispositions such as impulsivity, oppositionality, or extraversion.

The surprisingly low test–retest reliabilities of CS scores reported by Sultan et al. (2006) have devastating implications for use of the Rorschach in clinical and forensic settings. For instance, these findings indicate that if a parent in a custody evaluation is tested with the Rorschach by a psychologist, and then retested a few months later by a second psychologist, the test results are likely to be quite different even if both psychologists were proficient users of the test (Meyer & Viglione, 2008, p. 289). Because the large majority of Rorschach scores are likely to fluctuate substantially from one testing occasion to the next, it would be a serious error to use them to make long-term predictions (that is, more than 3 months) or to measure lasting personality characteristics. This deficiency of the Rorschach, aside from the test's other shortcomings, renders it useless for addressing most questions likely to arise in clinical and forensic evaluations such as, for example,

whether an examinee is likely to exhibit future problems with depression, interpersonal relationships, antisocial behavior, or parenting.

Validity

The literature on the validity of the Rorschach is so large that it is impossible to adequately review it within the constraints of this chapter. Interested readers should consult reviews, both supportive and critical, of this literature (Garb et al., 2001; Hunsley & Bailey, 1999, 2001; Lilienfeld, Wood, & Garb, 2000; Mihura, Meyer, Dumitrascu, & Bombel, 2013; Viglione & Meyer, 2008; Weiner, 1997; Wood, Lilienfeld, Garb, & Nezworski, 2000; Wood et al., 2003; Wood, Nezworski, Garb, & Lilienfeld, 2006).

A number of meta-analyses have been conducted on the Rorschach over the past 30 years. The earliest (Atkinson, 1986; Parker, 1983) used meta-analytic techniques that are now considered inappropriate. Later meta-analyses (Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neulieb, 1999; Parker, Hanson, & Hunsley, 1988), although providing some limited evidence for the validity of the Rorschach, have also been criticized on a number of methodological and statistical grounds (Garb, Florio, & Grove, 1998; Garb et al., 2001; Hunsley & Bailey, 1999; Mihura et al., 2013). Taken together, these studies provide at least some evidence for the validity of a small number of CS variables. But as both critics and proponents of the Rorschach agree, the validity of the Rorschach must be established for every scale (Hunsley & Bailey, 1999, 2001; Weiner, 2001; Wood, Nezworski, & Stejskal, 1996).

As the Rorschach is regarded as a test of personality and psychological functioning, it should correlate significantly with other measures of these qualities. Interestingly, this is not the case. Meta-analytic data indicate virtually no relation with other projective measures (a weighted mean *r* of .03; Hiller et al., 1999). If one considers the vast array of self-report measures of psychological functioning, the results are nearly as poor (a weighted mean *r* of .08; Mihura et al., 2013). However, in the face of hundreds of studies indicating weak associations between Rorschach scores and self-report indices purportedly assessing the same constructs, some Rorschach proponents have argued that one should expect such relations (e.g., Ganellen, 1996; Viglione, 1999). According to proponents (Mihura et al., 2013, the Rorschach can detect "implicit" characteristics that lie outside of awareness or that the examinee is unwilling to report, whereas self-report indices measure "explicit" characteristics that the examinee recognizes and is willing to acknowledge. Thus, the lack of association between Rorschach scores and self-report scales, according to the proponents, reflects the test's strength: a superior sensitivity to qualities that are unconscious or concealed.

One area where the Rorschach has demonstrated modest success is in the measurement of intelligence. Some Rorschach scores (e.g., Lambda,

Zf, DQ+) reflect the degree to which an examinee's responses to the inkblots are rich, complex, and well organized. These scores are intended to reflect the complexity of the examinee's thought processes, and research has shown that they are moderately correlated with scores on intelligence tests and tend to be lower in diagnostic groups with impaired thinking (i.e., mental retardation; Alzheimer's disease) than in other groups (Mihura et al., 2013). However, this positive finding is of limited practical use to psychologists because intelligence is much more accurately assessed by intelligence tests than by the Rorschach.

In addition, many investigators have examined the links between Rorschach responses and diagnostic measures. Wood et al. (2000) examined over 150 Rorschach studies that examined psychiatric diagnoses such as schizophrenia, depression, posttraumatic stress disorder, dissociative identity disorder, dependent personality disorder, narcissistic personality disorder, borderline personality disorder, antisocial personality disorder, and psychopathy. They found that diagnostic groups with psychotic or psychotic-like symptoms (schizophrenia, bipolar disorder, schizotypal personality disorder, borderline personality disorder) tend to give more deviant verbalizations (i.e., making odd or illogical statements while taking the test) and show more distorted form quality (i.e., providing responses that do not fit the shape of the inkblots) on the Rorschach than do other diagnostic groups or nonpatients. Because odd speech, disorganized thinking, and distorted perceptions are diagnostic criteria for most of these disorders, it is hardly surprising that these abnormalities are also expressed in Rorschach responses (for further evidence regarding the relationship of deviant verbalizations and distorted form to psychotic symptoms, see Jørgensen, Andersen, & Dam, 2000, 2001; Mihura et al., 2013).

Given the prominence accorded the Rorschach by many clinicians, Wood and colleagues' other conclusions are sobering. They found no strong evidence that the Rorschach can consistently detect depression, posttraumatic stress disorder, antisocial personality disorder, psychopathy, or any other psychiatric diagnoses. This conclusion is consistent with the results of the Hiller and colleagues (1999) meta-analysis, which found a weighted mean *r* of .18 when psychiatric diagnoses were used as the criterion in Rorschach validity studies (for further negative findings regarding the Rorschach and diagnoses, see Jørgensen et al., 2000, 2001; Wood et al., 2010; but for positive findings regarding the Rorschach and depression, see Mihura et al., 2013). Considering that the Rorschach has been touted for decades as a key measure of psychiatric disorders (e.g., Exner, 1993; Rapaport, Gill, & Shafer, 1946; Weiner, 1966), attempts by some Rorschach proponents to dismiss this mainly negative research evidence by arguing that the Rorschach was never intended to be a diagnostic test (e.g., Weiner, 1999) appear to be little more than post hoc rationalizations for consistently negative results.

Rorschach proponents have often suggested that the best way to

evaluate the validity of the Rorschach is to examine its incremental validity (e.g., Widiger & Schilling, 1980). However, a review by Garb (1984) concluded that the addition of Rorschach data to demographic or self-report personality data never increased the accuracy of clinicians' judgments when making personality assessments. It should be noted, however, that none of the studies included in Garb's review used the CS.

Given the limited evidence of convergent validity between the Rorschach and self-report measures, some Rorschach proponents have contended that this provides an opportunity for the Rorschach to add important clinical data beyond that available from such measures as the MMPI (e.g., Mihura et al., 2013; Weiner, 1993). To date, however, this argument has received only limited support from research that directly examined the incremental validity of CS scores. By far the most compelling findings on this topic were reported by Dao, Prevatt, and Horne (2008; see also Ritsher, 2004), who found that Rorschach scores, when added to MMPI-2 scores, provided incremental validity for the identification of schizophrenia spectrum disorders. Other studies have reported significant incremental validity for a few other Rorschach scores (for a review, see Meyer & Viglione, 2008), but none of these findings have been consistently replicated. Thus, except for Rorschach indicators related to schizophrenia and psychosis, the incremental validity of CS scores has yet to be clearly demonstrated.

The points discussed thus far regarding the validity of Rorschach scores are relatively noncontroversial. For instance, there is broad consensus among proponents and critics of the test that some Rorschach scores are related to psychotic thought processes and intelligence. There is also general agreement that Rorschach scores have shown little if any relation with psychiatric diagnoses (except those with a psychotic component), scores on other projective tests, or scores on self-report questionnaires. More controversial, however, are the conclusions presented in a recent meta-analysis by Mihura et al. (in press), which examined the relationship of CS scores with "external" criteria, which is to say, with variables that are measured by observational ratings, behavioral performance, or other objective evidence that does not depend on self-report.

Based on their meta-analytic findings, Mihura et al. (2013) concluded that several important CS scores lack demonstrated validity and that the Rorschach scores with the strongest research support are those that assess cognitive processes (e.g., psychotic thinking, distorted perception, intelligence). More controversial, however, were the authors' conclusions that some Rorschach scores are related to examinees' negative emotional experiences (e.g., distress, helplessness, agitation), as well as to interpersonal relationships, interpersonal perceptions, and desire for interpersonal closeness. And even more controversial were the authors' conclusions that the Rorschach has demonstrated validity as a measure of emotional impulsivity and suicide risk, characteristics that can play an important role in clinical and forensic decision making. There is no doubt that many of the Mihura et

al. conclusions will be carefully scrutinized by both proponents and critics of the Rorschach, and that much more debate and research will be necessary before any consensus can emerge regarding many of these authors' conclusions.

Conclusions

Of the tests we review in this chapter, the Rorschach has the longest history of empirical investigation. Exner's CS brought some order to the field and inspired a more scientific approach to the Rorschach. Nevertheless, many optimistic claims about the test have failed to stand up to scrutiny. For instance, contrary to Exner's claim that the Rorschach can be administered by almost anyone with some CS training, research has shown that even examiners with solid CS training are likely to make serious errors administering the CS, and that these errors are undetectable to other professionals who examine their work. Although Exner's norms were widely praised for decades, eventually they were found to be seriously in error, with a tendency to make psychologically healthy children and adults appear psychologically disturbed. Similarly, optimistic claims regarding the test–retest reliability of CS scores have been shown to be misleading: Most CS scores have low test–retest reliability and are unsuitable for measuring stable personality traits or long-term behavioral dispositions.

Some Rorschach scores have a demonstrated valid relationship with psychotic thought processes and intelligence. Otherwise, there is little evidence that the Rorschach possesses adequate convergent validity when examined vis-à-vis data from other projective tests, self-report measures of personality, or diagnostic categories. Despite decades of research, there has been no convincing accumulation of data supporting the use of the Rorschach in routine clinical practice, although a recent meta-analysis by Mihura et al. (2013) has reopened the debate about Rorschach validity and utility. Until the controversies over the Rorschach are settled, contemporary critics of the Rorschach will probably continue to conclude that there is insufficient scientific evidence to justify the continued use of the test in clinical settings (Garb, 1999; Hunsley & Bailey, 1999, 2001).

Thematic Apperception Test

The Thematic Apperception Test (TAT; Murray, 1943) is a projective technique in which examinees are asked to tell stories about pictures printed on cards. According to its developer, in creating these stories examinees reveal dominant drives, emotions, and conflicts of their personality, some of which may not be consciously accessible. The TAT consists of 31 cards, with some cards intended for administration to male examinees and others intended for female examinees. Research does not support the clinical utility of the gender-specific card sets, as examinee gender and sex role

are unrelated to the stories given to gender-matched and unmatched cards (Katz, Russ, & Overholser, 1993). Following the development of the TAT, other "storytelling" tests were developed for children, older adults, and ethnic minority groups (Teglasi, 2010). For many decades the TAT and similar tests were commonly used by clinical psychologists and, at least until the end of the 1990s, the proportion of users remained relatively constant (Camara et al., 2000; Watkins et al., 1995). There are no data available on TAT use in recent years, but a decade ago it was still included in many graduate courses in clinical assessment (Childs & Eyde, 2002) and many internship training directors recommended that trainees have experience with the TAT (Clemence & Handler, 2001). It is worth noting, however, that diplomates in forensic psychology are unlikely to use the TAT and that most view it as unacceptable for use in any type of forensic evaluation (Lally, 2003).

Standardization and Norms

Little information is available on the extent to which practitioners follow or modify the original TAT instructions developed by Murray (1943). There is no consistency in the number of cards selected, the specific cards administered, and the order in which they are presented (Groth-Marnat, 2009). Variability in card selection and administration renders comparisons across the research literature virtually impossible and discussions of reliability and validity largely futile (Keiser & Prather, 1990; Vane, 1981).

In addition to Murray's scoring and interpretation (1943), a number of different systems have been developed for the TAT (see Jenkins, 2008, for a presentation of many of these systems). The earliest efforts to establish the TAT as a scientifically sound test were made by Eron (1950, 1953) and Murstein (1965, 1972), who developed normative data and a list of themes typically obtained for each card. Using some of the TAT cards and adding other picture stimuli, McClelland and colleagues conducted a programmatic series of studies on achievement, power, and affiliative needs (for reviews, see McClelland, Koestner, & Weinberger, 1989; Spangler, 1992). The Social Cognition and Object Relations Scales (SCORS) scoring system has some empirical support for assessing aspects of interpersonal relations (e.g., Westen, 1991). However, the necessary stimuli and scoring protocols for these approaches have not been routinely incorporated into clinical practice. Indeed, among clinicians, there is no consensus regarding which systems should be used, and there is no comprehensive set of norms that can be used in scoring and interpreting the TAT (Teglasi, 2010).

Even ardent proponents of the TAT recognize that most clinicians do not use research-based scoring and interpretative systems; thus, these proponents have abandoned a psychometric approach to the test (Dana, 1985; Pinkerman, Haynes, & Keiser, 1993; Rossini & Moretti, 1997). The TAT is therefore best characterized as a technique that (1) is taught and used in

a manner that ignores scientific and professional standards and (2) emphasizes the clinician's intuitive interpretative skills (Rossini & Moretti, 1997). The problems with such a stance are clear and have been raised repeatedly by TAT proponents and critics alike (e.g., Garb, 1998; Holt, 1999; Lilienfeld et al., 2000).

Reliability and Validity

Given typical usage by clinical psychologists, there are no data supporting the reliability or validity of the test. Many of the scoring systems developed for use with TAT stimuli lack evidence of reliability and validity. For example, for many systems reviewed in a recent TAT handbook (Jenkins, 2008), (1) reliability is based on indices of agreement between raters, which is an unacceptable estimate of interrater reliability that does not make adjustments for chance agreement, and (2) reports of validity are often sparse, dated, not replicated, limited to specific elements of a scoring system instead of the whole system, or of questionable clinical relevance. In a recent evaluation of the TAT research literature, McGrath and Carroll (2012) recommended that a minimum of 10 cards be used to have the chance of attaining a reasonable level of reliability—few scoring systems actually use this many cards. As a result of all of these problems, even strong proponents of the TAT such as Jenkins (2008) admit that it is, at best, a future possibility that research-based scoring systems might have a widespread clinical impact.

Cramer (1991) has developed a scoring system for assessing defense mechanisms from TAT stories by providing clear instructions for card selection, administration, and scoring. Although interrater reliabilities for the measures of denial, projection, and identification were acceptable, internal consistency, alternate forms reliability, and retest reliability were unacceptably low (Cramer, 1991). Cramer (1999) reviewed over a dozen studies suggesting that the scoring system had demonstrated validity in a range of populations, including children, adolescents, nonpatient adults, and adults with psychiatric conditions. However, closer examination of these results reveals little consistency across studies in how the defense mechanism scores are related to measures of adjustment and distress—an issue of particular importance for any clinical use of the system. Most recently, Cramer has applied her system to the study of the longitudinal development of defense mechanisms (e.g., Cramer, 2009).

As previously mentioned, Westen developed the SCORS, a psychodynamically oriented scoring system for TAT responses. His system focuses on the assessment of object relations (complexity of representations of people, affect tone of relationship schemas, capacity for emotional investment in relationships, and understanding of social causality). Using a detailed scoring manual and data from five to seven TAT cards, high interrater reliability has been obtained in several studies for the SCORS (e.g., Westen,

Lohr, Silk, Gold, & Kerber, 1990). Data from nonpatient and patient samples indicate convergent validity with a range of self-report, interview, and projective measures (see Westen, 1991). Studies have found that the system can differentiate between clinical and nonclinical samples, including children with and without a history of physical abuse (Freedenfeld, Ornduff, & Kelsey, 1995), and among individuals with personality disorders (Ackerman, Clemence, Weatherill, & Hilsenroth, 1999).

TAT scoring systems have also been developed from nonpsychodynamic theoretical orientations. Peterson and Ulrey (1994), for example, developed a coding scheme to evaluate attributional style. Using written responses to four TAT cards chosen for negative content, they found that attributional dimensions of stability and globality could be reliably rated and that the globality index correlated significantly with a self-report measure of attributional style. Ronan and colleagues developed a scoring system to tap personal problem-solving strategy (see Ronan, Gibbs, Dreer, & Lombardo, 2008). In this system, three TAT cards are used, with evidence of good reliability and validity. For example, in randomized control studies, participants who received training in decision-making and problem-solving skills scored higher on the TAT problem-solving index than did participants who received only minimal exposure to these skills.

Conclusions

Several lines of research have demonstrated that responses to TAT and similar stimuli can yield reliable and valid information about focused aspects of psychological functioning. Nevertheless, these encouraging findings are almost totally irrelevant to the clinical use of the TAT, as most clinicians use the test as a global measure and do not use standardized scoring systems that have some empirical support. Of concern, TAT proponents frequently and inappropriately overgeneralize from the results of this research to clinical applications of the test, suggesting that the TAT and other projective tests are valid and scientifically supported (e.g., Ackerman, Fowler, & Clemence, 2008; Masling, 1997). Such claims are unfounded and largely misleading. A careful examination of the literature reveals that, as typically used in clinical practice, the TAT is a potentially useful measure that falls well short of professional and scientific test standards (cf. McGrath & Carroll, 2012).

Projective Drawings

Various projective drawing techniques require the examinee to draw a person (Draw-A-Person [DAP]; Harris, 1963; Human Figure Drawing [HFD]; Koppitz, 1968), a house, a tree, and a person (House–Tree–Person [HTP]; Buck, 1948), or a family engaged in some joint activity (Kinetic Family Drawing [KFD]; Burns & Kaufman, 1970). Drawings are used by mental

health practitioners of various disciplines, including art therapists, school counselors, psychiatrists, and psychologists, to assess intellectual or emotional functioning in adults and children.

Proponents view drawings as providing access to unconscious material that the person would otherwise be unable or unwilling to communicate (Handler, 1985). They consider projective drawings to be especially useful in the assessment of persons who are guarded or lack verbal skills. Proponents suggest that drawings can provide a springboard for discussion, as well as help assess progress in psychotherapy. We restrict our review here to projective drawings and do not address the use of human figure drawings as an aid to children's recall of touching (e.g., Aldridge et al., 2004; Bruck, 2009).

An abundant literature describes the presumed importance of different characteristics of the drawings, including size, the inclusion or absence of various features, the heaviness of the lines, the relative distance of figures from one another, and the complexity of the drawing. Machover (1949) proposed links between features of the drawings and various psychological features; for example, large eyes were assumed to reflect paranoia, heavy shading aggressive impulses, and repeated erasures anxiety. Efforts to integrate clinical lore surrounding the meaningfulness of diverse indicators have resulted in the publication of numerous scoring systems. Koppitz (1968) was the first to promote interpretation based on the whole drawing rather than on single indicators of pathology. Applying Koppitz's integrative principles, Naglieri and Pfeiffer (1992) developed the Draw-A-Person Screening Procedure for Emotional Disturbance (DAP-SPED), which shows some promise in distinguishing clinical from nonclinical samples.

We know of no data on the numbers of practitioners who apply various scoring systems. Thus we are again faced with the dilemma of determining the pertinence of reliability and validity studies for actual clinical practice. Of great concern are articles advocating the use of drawings to assess indications of sexual abuse in children's artwork (e.g., Riordan & Verdel, 1991). Proponents of projective drawings clearly attach great weight to their interpretations and attribute special powers to certain clinicians judged to possess particular acumen in interpreting drawings.

Standardization and Norms

There is variability in the instructions given to examinees completing projective drawings. Typically, examinees are provided with a supply of blank sheets of paper, pencils, and erasers. Instructions specify the size of the paper provided, but these instructions vary across protocols. As children's drawings are influenced by the materials that are provided (Burkitt & Barrett, 2011), these instructions may affect the drawings produced. In the DAP, respondents are instructed to draw a picture of a person (Handler, 1985). The HTP requires the person to draw as good a house, tree, and

person as possible (Hammer, 1985). The KFD requires the child to draw his or her family doing something. Directions are kept to a minimum, although some instructions explicitly state that the drawing should be of a person rather than a stick figure or cartoon character. In some approaches, examinees are asked to create a story about the drawing or are asked about their associations to the drawing, such as whether the person depicted is happy or sad.

Given the large number of different projective drawing techniques and myriad scoring systems, it is not surprising that there are inadequate norms (Handler & Habenicht, 1994). Although a number of scoring systems have been developed, considerable use is made of idiosyncratic scoring, thus rendering norm-based interpretation impossible. One important issue that has not been sufficiently addressed is the examinee's level of competence in drawing or willingness to draw (Feldman & Hunt, 1958; Nichols & Strumpfer, 1962). Given the important gender and age differences in children's drawings (Deaver, 2009), before projective techniques could be used as part of a scientifically based assessment, it would be necessary to develop knowledge of the characteristics of drawings produced by examinees across ages, genders, and levels of aptitude and interest in drawing.

Reliability and Validity

Scoring the qualities of drawings produced over a short interval has been shown to have some reliability, although reliability between raters tends to be poor (Palmer et al., 2000; ter Laak, de Goede, Aleva, & van Rijswijk, 2005; Vass, 1998). As a given feature yields different clinical hypotheses, however, reliability at the level of the overall interpretation is reduced (Thomas & Jolley, 1998). For example, in West's (1998) meta-analysis of the validity of projective techniques in discriminating abused and nonabused children, variations in head size were seen as indicative of sexual abuse in one study and of physical abuse in another.

Evaluating the validity of projective drawings is rendered more difficult by the variety of techniques and scoring systems. An intense and often acrimonious discussion has arisen regarding the validity of projective drawings (e.g., Holtzman, 1993; Joiner, Schmidt, & Barnett, 1996; Motta, Little, & Tobin, 1993; Riethmiller & Handler, 1997). Critics note the lack of evidence for their validity (Joiner et al., 1996; Motta et al., 1993), whereas proponents castigate critics for their failure to recognize the rich clinical material that is accessible to nonskeptics (Reithmiller & Handler, 1997). Notably, there is little evidence that projective drawing "experts" are accurate in their capacity to identify different types of psychopathology (Wanderer, 1997).

An emerging consensus is that data do not support the use of single indicators or signs of pathology, but that the drawings must be interpreted in a holistic fashion within the context of other assessment data (Flanagan

& Motta, 2007; Handler & Habenicht, 1994; Tharinger & Stark, 1990). As an example, Matto (2002) found encouraging evidence of concurrent validity in the DAP-SPED, in that drawing-related scores accounted for variance in a measure of parent-reported internalizing problems, over and above the contribution of another parent-rating of children's internalizing problems. Preliminary findings also suggest equivalence of composite scores across white–black and white–Hispanic pairs matched on socioeconomic variables (Matto & Naglieri, 2005).

In an effort to circumvent the issue of drawing ability, Williams, Wiener, and MacMillan (2005) asked three groups of children to build human figures choosing prefabricated pieces of human figure body parts. Some of the components represented features that have been hypothesized to be indicators of sexual abuse. These authors found no evidence that any of the indicators singly or in combination provide evidence of sexual abuse.

A fundamental assumption of projective techniques is that they provide access to the unconscious and that, by communicating with the client through a more "primitive" channel, the clinician may discover the client's true feelings and experiences. Thus, drawings are assumed to circumvent the client's defenses and provide invaluable material about psychological functioning that the client might otherwise be unable or unwilling to acknowledge. This assumption does not meet the scientific criterion of falsifiability because the criterion against which the technique should be assessed is not measurable through other means.

Conclusions

Projective drawings maintain their place in the armamentarium of many psychologists despite the paucity of scientific evidence for their usefulness. Currently, diverse techniques with various scoring systems are being used to discover aspects of functioning that the examinee is assumed to be unable or unwilling to express directly. Until hypotheses based on projective drawings are formulated in a manner that can be subjected to scientific scrutiny and are supported in rigorous studies, there can be no basis to the claims for the validity of these approaches. Projective drawings do not meet standards for admissibility of evidence in a court of law (Lally, 2001). Given advances in other methods of assessing intelligence and psychological functioning, the investment of resources in the use of projective techniques that are vulnerable to a host of weaknesses and errors in administration, scoring, and interpretation does not represent a reasonable cost-benefit ratio (cf., McGrath & Carroll, 2012).

Anatomically Detailed Dolls

Clinicians working with young children have used dolls and puppets on the assumption that children can express material through this medium that

they are unable or unwilling to express verbally (Sattler, 1998). Children's limited cognitive and language abilities are especially challenging in the context of sexual abuse allegations because the alleged child victim may be the sole witness to the offense and clinicians must ensure that investigative interviews do not further traumatize the child (Cross, Jones, Walsh, Simone, & Kolko, 2007). Health professionals and law enforcement personnel working in the field of child sexual abuse have turned to anatomical dolls with features such as anal and vaginal orifices, penises, pubic hair, and breasts as a possible solution to the challenges of assessing child victims (Conte, Sorenson, Fogarty, & Rosa, 1991).

A strategy that holds the promise of obtaining accurate information in a child-friendly manner to secure the conviction of a perpetrator of child sexual abuse is appealing. However, research has raised serious doubts about the usefulness of anatomical dolls in generating additional accurate information in the assessment of young children (e.g., Ceci & Bruck, 1995; Thierry, Lamb, Orbach, & Pipe, 2005). Nevertheless, anatomical dolls are regarded by some as an essential tool in child protection efforts (e.g., Faller, 2005) and may be accepted in testimony in child abuse cases in 10 states (National District Attorneys Association, 2010). Moreover, an analysis of 500 videotaped forensic interviews in cases of suspected child sexual abuse found that anatomical dolls were introduced in 49% of interviews, with interviewees ranging in age from preschool to adolescence (Hlavka, Olinger, & Lashley, 2010).

Standardization and Norms

Anatomical dolls are produced by several manufacturers and are available with different facial expressions, facial features, and racial physiques. Dolls may be clothed, semi-clothed, or naked. Numerous protocols describe the ways that dolls may be used in interviews (Everson & Boat, 1994; Hlavka et al., 2010). In some protocols, the child is instructed not to play with the doll. When the child's behavior with the doll demonstrates fantastic details, such as stating the doll is flying (e.g., Thierry et al., 2005), the interviewer can easily recognize play, but at other times it may be difficult to judge whether the child has understood and followed the directive not to play with the doll. The stage in the investigation when dolls are introduced varies across protocols (Thierry et al., 2005), with some indicating that they should be presented to the child only after a disclosure of abuse (e.g., Hlavka et al., 2010) and others suggesting that this defeats the purpose of using the dolls in the case of children who are unable or unwilling to disclose abuse (e.g., Faller, 2005).

Among the many uses of the dolls, they are employed as *anatomical aids* when the child is asked to name body parts; as *demonstration aids* when the interviewer requests that the child show with the doll what happened; and as a *memory stimulus/diagnostic screen* when the child is left

to play with the doll (either in the presence of the examiner or alone, but unobtrusively observed) and behavior with the dolls is presumed to reflect experiences (Everson & Boat, 1994; Faller, 2005). Interpretation of the child's behavior with the dolls requires a solid research base on the range of behaviors with anatomical dolls typical of children who have and who have not been abused. Given the lack of norms on children's behavior with the dolls, experts agree that decisions about whether a child has been abused should not be made solely on the basis of their behavior with dolls (Dickinson, Poole, & Bruck, 2005; Everson & Boat, 1994; Koocher et al., 1995; Pipe & Salmon, 2009).

Reliability and Validity

Given the lack of standardization in the ways anatomical dolls are used, it is not surprising that little research has addressed the reliability of assessments in which they are employed. The use of anatomical dolls requires that children have representational insight, meaning that they understand that the doll represents their bodies (Dickinson et al., 2005). Some protocols are designed to assess this understanding by asking the child to select a doll that is most like him/herself and another doll that is most like the alleged perpetrator. Although some kind of matching is necessary for representational insight, it may not be sufficient to ensure that the child has grasped the concept that the doll represents him/herself (Dickinson et al., 2005). Similarly, exhortations to the child not to use the doll for play (e.g., Hlavka et al., 2010) may be of limited usefulness in very young children for whom the boundaries between play and nonplay are unclear.

For dolls to be used as a demonstration aid, it is necessary to have evidence that the dolls help children to provide additional accurate information beyond the information that is verbally disclosed. Thierry and colleagues (2005) reviewed videotaped interviews in which anatomical dolls were introduced as demonstration aids at various stages in the investigation of child abuse. These researchers found that the use of dolls did not enhance the quality of information obtained from young children. In fact, several children reported fantastic details, suggesting they were using the dolls to play. Evidence suggests that dolls enable some older children to recall more details of the abuse, but Thierry and colleagues cautioned that the questionable accuracy of these recollections outweighs their memory aid benefits.

The American Psychological Association (1991) commissioned two task forces to examine the validity of data based on use of anatomically detailed dolls. The first concluded that, although the dolls are not standardized and although there are no normative data and no uniform standards for conducting interviews, doll-centered assessment "may be the best available practical solution for a pressing and frequent clinical problem (i.e., the investigation of the possible presence of sexual abuse of a child)"

(American Psychological Association, 1991, p. 722). Furthermore, the task force exhorted psychologists who undertake doll-centered assessment to be competent (although this competency was not defined) to document their procedures and to provide clinical and empirical rationales for their procedures and interpretations. These recommendations reflect a puzzling mixture of reliance on unspecified clinical wisdom coupled with reference to a research literature that the task force had concluded was nonexistent.

The Anatomical Doll Working Group, funded by the American Psychological Association (Koocher et al., 1995), reiterated the conclusion of the earlier task force (American Psychological Association, 1991) that anatomical dolls do not meet any of the criteria for a valid psychological test or a projective technique. Koocher and colleagues (1995) advised that conclusions about child sexual abuse cannot be made on the basis of doll play alone and that reports of children under 4 years of age are particularly prone to be affected by misleading questions. These cautions notwithstanding, Koocher and colleagues reasserted the original American Psychological Association task force position. Both of the Association's resolutions reflect the tensions in psychological practice and the lip service paid to science by some psychologists who are willing to examine research literature but equally willing to dismiss it if it does not correspond to views founded on clinical experience.

In reviewing the literature, proponents and critics differ in their interpretation of the research findings and in their willingness to tolerate Type I and Type II errors. Advocates focus on possible increases in the reporting of sensitive material with anatomical dolls (e.g., Everson & Boat, 1994; Faller, 2005); argue that anatomical dolls are no worse than other assessment strategies (Koocher et al., 1995); and lament that research has not kept pace with practice (Hlavka et al., 2010). This stance stands in sharp contrast to the prototypical scientific position we noted earlier, namely, that the onus of demonstrating the utility of a particular assessment strategy rests with its proponents.

Critics focus on possible errors and inaccuracies in reporting that are associated with the use of anatomical dolls (Dickinson et al., 2005; Thierry et al., 2005) and charge that the incremental validity of anatomical dolls has not been demonstrated (e.g., Ceci & Bruck, 1995; Dickinson et al., 2005; Thierry et al., 2005; Wolfner, Faust, & Dawes, 1993). That is, they argue that anatomical dolls must be shown to consistently *add* to our ability to determine whether a child has been abused above and beyond already available information, such as interviews, observations, and rating scales. Hlavka and colleagues (2010) recently concluded that the high concordance between interviewers' stated reasons for using the dolls and interviewers' perceptions of the value of the dolls for that purpose "demonstrates valid and reliable assessments" (p. 537). This reasoning is a troubling example of mistaking user perceptions for scientific validity (i.e., if people use it and think it is useful, it must be valid). Equally worrisome are comments

suggesting that attacks on the use of anatomical dolls reflect efforts to obscure the prevalence of sexual abuse of children (e.g., Faller, 2005). The controversies surrounding usefulness and suitability notwithstanding, the use of anatomical dolls in forensic interviewing is included in the curriculum of the week-long course offered by the American Prosecutors Research Institute (American Prosecutors Research Institute, 2003).

Conclusions

Although it would be very helpful to have a tool that yields accurate information from young children who are cognitively or emotionally unable to explain clearly what happened, the use of anatomical dolls requires sophisticated cognitive processing that they may not yet have developed. Research indicates that we may have greatest confidence that older children (who can also respond verbally) may recall more details when dolls are used as a memory aid, but this must be balanced against concerns about the accuracy of these additional recollections. Many proponents of anatomical dolls adopt scientific language by referring to "evidence," "studies," "research," and "empirical support." However, they seek to absolve the procedure from the scrutiny of scientific standards by denying that anatomical dolls constitute a psychological test. Paradoxically, some proponents promote their approaches as scientifically supported while rejecting arguments that these measures be held to scientific standards. We reiterate previous findings that neither the stimuli nor the procedures used in anatomical doll assessments are standardized. We reject claims that anatomical dolls can be used as screening instruments without meeting the standards for psychological tests, and therefore we strongly advise against their use for this purpose in investigations of child sexual abuse.

Myers–Briggs Type Indicator

The Myers–Briggs Type Indicator (MBTI; Briggs & Myers, 1998) is a self-report test based on Jung's theory of personality types. The four basic personality preferences are operationalized in the MBTI as bipolar, continuous constructs: extraversion–intraversion (oriented outwardly or inwardly), sensing–intuition (reliance on sensorial information versus intuition), thinking–feeling (tendency to make judgments based on logical analysis or personal values), and judgment–perception (preference for using either thinking–feeling or sensing–intuition processes for interacting with the world). Based on scores on these four dimensions and their relation to established cutoff scores, examinees are assigned to 1 of 16 personality-type categories (e.g., extraverted, sensing, thinking, judgment). Use of these 16 categories is controversial, as they are not consistent with Jungian theory and are not supported by data gathered from the MBTI (Barbuto, 1997; Garden, 1991; Girelli & Stake, 1993; Pittenger, 1993, 2005). We address this point in more detail later in this chapter.

The current versions of the MBTI are Form M, a forced-choice 93-item test, and Form Q, a forced-choice 144-item test. These forms have been translated and normed in many languages and are reportedly among the most commonly used measures of normal personality (Schaubhut, Herk, & Thompson, 2009; Schaubhut & Thompson, 2011). Although it was developed for use in education, counseling/therapy, career guidance, and workplace team-building, it is within the assessment practices of career guidance and personnel selection that the MBTI has gained dominance—so much so that it is routinely used by psychologists for gathering information on possible career paths and job placement (Jackson, Parker, & Dipboye, 1996; McCaulley & Martin, 1995; Turcotte, 1994). The research on the MBTI is impressive in scope, with hundreds of studies published in the past several decades on a range of personality, educational, and vocational constructs.

Standardization

As the MBTI is a published self-report test, standardization of test instructions and test items is assured when the test is administered appropriately. The MBTI manual (Briggs & Myers, 1998) provides instructions on scoring and converting the scores to 1 of the 16 personality types. Information is also provided on how the results of the test should be interpreted, both in general terms and with reference to counseling, educational, and career counseling contexts. A central concern with the use of the 16 types raised repeatedly in the research literature is the appropriateness of the cutoffs used to assign examinees to a type. Researchers have found that scale scores close to the cutoffs frequently lead to classification errors, resulting in repeated calls for alterations in scoring procedures (e.g., Girelli & Stake, 1993; Harvey, Murry, & Stamoulis, 1995; Tzeng, Ware, & Chen, 1989).

Norms

The manuals contain normative information based on data from thousands of men and women of different ages and occupations. Research has tended to support the appropriateness of the norms associated with many previous versions of the MBTI and to find them applicable across minority groups and cultures (e.g., Kaufman, Kaufman, & McLean, 1993).

Reliability and Validity

The MBTI preference scores have acceptable levels of internal consistency and test–retest reliability (Carlson, 1985; Schaubhut et al., 2009); however, far less evidence is available on the reliability of the 16 types. This is a critical issue, as all interpretations are based on the test-taker's type, not on the scores on the continuous variables. As Matthews (2004) nicely put it, the use of type scores is akin to measuring people's height (a normally distributed, continuous variable) and then categorizing them as "short" or

"tall" using the mean population height as the cutoff score. The scoring and reliability of the 16 types present a host of problems (Matthews, 2004; Pittenger, 2005): (1) There is no evidence that scores on the preference constructs are bimodal or bitangential, (2) the use of a type system means that there can be greater preference score differences between two test-takers receiving the same type classification than between two who obtain preference scores close to, but on either side of, the cutoff score, and (3) reliability indices for the preference score, as reported in the supplemental manual (Schaubhut et al., 2009), are essentially irrelevant in gauging the reliability of the 16 types.

The scientific literature includes numerous validity studies that relate MBTI preference scores and types to myriad personality constructs, ability measures, and occupations. Nevertheless, no attempt has been made to integrate the data from these studies to guide the valid interpretation of the test results. Moreover, there is relatively limited information on the predictive validity (i.e., whether accurate forecasts of educational and career choices can be made on the basis of the MBTI types) or the incremental validity and utility of the MBTI (i.e., whether the MBTI types meaningfully add to the prediction of these decisions; whether there are optimal educational, career, or employment decisions made on the basis of MBTI data).

One aspect of the MBTI that has received extensive attention in the literature is the validity of the test as a measure of personality. Using a range of analytical procedures, including exploratory factor analysis, confirmatory factor analysis, and cluster analysis, researchers have generally found that (1) the observed factor structure of the MBTI is consistent with the hypothesized four personality preferences (Thompson & Borrello, 1986; Tischler, 1994), although often at a less than optimal level (Harvey et al., 1995; Jackson et al., 1996; Sipps, Alexander, & Friedt, 1985), and (2) the fit between the hypothesized 16 types and actual test data is poor (Lorr, 1991; Pittenger, 1993; but see Pearman & Fleenor, 1996).

As a measure of global personality, the MBTI has been criticized for its failure to relate to other well-established vocational and personality measures. Although the efforts of the test developers to include concurrent validity data with a range of such measures in the test manual is laudable, little consistent evidence has been found that the four personality preferences relate to comparable constructs assessed by other measures. Published research suggests that the MBTI bears little correspondence to measures of vocational preferences and job performance (e.g., Apostal & Marks, 1990; Furnham & Stringfield, 1993). The four preference scores of the MBTI have limited correspondence with prevailing, well-validated models of personality structure (e.g., Furnham, 1996; Furnham & Crump, 2007; Furnham, Dissou, Sloan, & Chamorro-Premuzic, 2007; McCrae & Costa, 1989; but see MacDonald, Anderson, Tsagarakis, & Holland, 1994). As with reliability evidence, findings of qualified support for the preference

scores have little relevance in determining the validity of the MBTI types. One can only conclude that the MBTI, especially when scored for the 16 types, is insufficient as a contemporary measure of personality.

Conclusions

The MBTI is based on an explicit theory of personality. It was developed and normed in a manner consistent with current standards, and it has been found to be reliable at the level of the four personality preferences. However, questions about the reliability and validity of the 16 personality types and evidence of limited correspondence between the MBTI and other global measures of personality and vocational interests render the test suspect as an assessment tool. Accordingly, psychologists are advised to rely on personality and vocational interest tests (e.g., the Strong-Campbell Interest Inventory) that have a stronger empirical basis (cf. Boyle, 1995).

CONCLUSIONS AND RECOMMENDATIONS

Psychologists face a daunting task in making sense of the vast literature on psychological assessment. In considering ways to address pressing clinical questions, they have access to a panoply of potential tools. Nevertheless, there are no simple ways to determine whether a test is scientifically valid. The fact that it is marketed in a prestigious professional newsletter or described in a scholarly journal provides no guarantee that it meets adequate test standards. For example, a published meta-analysis (West, 1998) provided data that appeared to support the use of projective techniques in detecting sexual abuse in children. However, reexamination of the data revealed that West had included in her calculations of effect sizes only significant findings, thereby inflating the apparent power of projectives to identify abused children (Garb, Wood, & Nezworski, 2000). Simply referring to research citations may therefore be inadequate in determining whether a test is appropriate, and many of the articles we reviewed contained abstracts claiming support for an assessment approach, although the articles offered at best mixed support.

Knowledge of psychometric principles and procedure is essential for psychologists involved in assessment activities. Numerous sources provide information on these issues (e.g., Haynes et al., 2011; Wood et al., 2006). We urge psychologists to use as a reference the most recent edition of the *Standards for Educational and Psychological Testing* (i.e., American Educational Research Association et al., 1999), which explains the principles that have guided a scientifically based approach to assessment.

Psychologists are required to engage in reasoned decision making as they select assessment tools. Knowledge of an assessment procedure may become obsolete as it is replaced by a more sophisticated understanding

of its limitations. Psychologists using assessment procedures that lack a published manual are required to conduct their own scholarly review of the pertinent literature to determine whether the test meets basic standards of reliability and validity, and whether suitable norms exist. Those using published materials are also required to familiarize themselves with the data relevant to use of the test. The fact that a published manual exists does not guarantee either that the test meets standards of reliability and validity or that appropriate norms are available. Professional and ethical standards indicate that each psychologist is responsible for determining (1) the question that the psychological assessment is designed to answer, (2) whether there is a test that is adequately standardized and that yields reliable and valid information, and (3) whether pertinent norms allow interpretation of the responses on the test in a given circumstance. We caution psychologists against uncritically accepting the argument that a given assessment procedure is absolved from the obligation to meet accepted standards.

Among the tests we reviewed, we found scant support for most Rorschach scores, a few promising avenues for the TAT (although no support for this measure as it is currently used in clinical practice), only very limited promise for holistic scoring of some projective drawings, no support for anatomical dolls as a screening instrument for evidence of sexual abuse, and evidence that the MBTI is a self-report measure that lacks convincing reliability and validity data for types derived from the test. A lack of standardization in the use of many of these techniques and an overreliance on unsubstantiated beliefs that certain people possess special interpretive powers have thwarted the possibility of advancing these techniques into the realm of scientifically supported assessment strategies.

Responsibility for demonstrating the adequacy of an assessment procedure rests, first, with those developing the procedure and, second, with psychologists electing to use it. Proponents of specific assessment procedures are responsible for elucidating a standardized protocol that fully explains administration procedure. In the case of the Rorschach, TAT, projective drawings, and anatomical dolls, it is necessary that proponents of each approach reach consensus on standardized administration and scoring. The maintenance of idiosyncratic versions of these tests obstructs their being established as scientifically sound assessment strategies.

Once a standard protocol is established, research must address issues of reliability. Decision rules must permit independent raters to reach the same judgments concerning a response. Once a tool can be administered in a standardized fashion and the examinee's responses judged consistently, the issue of validity can be addressed. Tests that purport to measure a construct that cannot be independently measured by other tests or other assessment techniques are inherently unfalsifiable and therefore unscientific (see also Lilienfeld, Lynn, & Lohr, Chapter 1, this volume). Finally, norms must be developed so that scores for an individual can be meaningfully interpreted. If clinical psychology is to be a fully scientific discipline, it is

essential that established criteria for psychological tests be applied consistently and demanded by researchers and practitioners alike.

GLOSSARY

Clinical utility: The extent to which the assessment makes a meaningful, desirable difference in clinical decision making, treatment planning, and/or treatment outcome.

Falsifiability: The extent to which a hypothesis, measure, or theory be framed so that it can be subjected to empirical investigation and potentially refuted by such an investigation.

Norms: Normative data obtained from a comparison group of similar individuals that, ideally, are representative of the population to which the comparison group belongs. Interpretation of the obtained test results requires a comparison between the examinee's test data and relevant norms.

Projective tests: Either ambiguous test stimuli and/or relatively unstructured tasks that, putatively, require examinees to structure their test responses in a manner that reveals basic, unconscious personality characteristics.

Psychological tests: The measurement of a sample of behavior obtained under standardized conditions and the use of established rules for scoring or interpreting the behavioral sample. Additionally, whenever a claim is made or implied that the accuracy or validity of assessment-based inferences stems from the way in which the sample was collected (i.e., the nature of the stimuli, technique, or process that gave rise to the sample of behavior), not just from the expertise, authority, or special qualifications of the assessor, then the process used to collect and interpret the behavioral sample should be considered a psychological test.

Reliability: A test's consistency, including whether (1) all aspects of the test contribute in a meaningful way to the data obtained (internal consistency), (2) similar results would be obtained if the test was conducted and/or scored by another evaluator (interrater reliability), and (3) similar results would be obtained if the person was retested at some point after the initial test (retest reliability or test stability).

Standardization: Necessary to ensure that the influence of unique aspects of the testing situation and the assessor are minimized and involves the provision of (1) comparable test stimuli across assessors, (2) detailed instructions regarding the administrative procedures, and (3) detailed descriptions of scoring procedures for obtained test data.

Type I and Type II errors: A Type I error involves accepting a hypothesis when the data do not support the hypothesis; a Type II error involves rejecting a hypothesis when the data are supportive of the hypothesis.

Validity: The extent to which the test measures what it purports to measure, including whether (1) the test samples the type of behavior that is relevant to the purpose of the test (content validity), (2) it provides data consistent with theoretical postulates associated with the phenomenon being assessed (concurrent and predictive validity), (3) it provides a relatively pure measure of the phenomenon that is minimally contaminated by other psychological characteristics (discriminant validity), and (4) it adds to our assessment-based knowledge above and beyond other information gathered in an assessment (incremental validity).

REFERENCES

Ackerman, S. J., Clemence, A. J., Weatherill, R., & Hilsenroth, M. J. (1999). Use of the TAT in the assessment of DSM-IV Cluster B Personality Disorders. *Journal of Personality Assessment, 73,* 422–448.

Ackerman, S. J., Fowler, J. C., & Clemence, A. J. (2008). TAT and other performance-based assessment techniques. In R. P. Archer & S. R. Smith (Eds.), *Personality assessment* (pp. 337–378). New York: Routledge/Taylor.

Acklin, M. W. (1995). Integrative Rorschach interpretation. *Journal of Personality Assessment, 64,* 235–238.

Acklin, M. W., McDowell, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *Journal of Personality Assessment, 74,* 15–47.

Alcock, J. E. (1991). On the importance of methodological skepticism. *New Ideas in Psychology, 9,* 151–155.

Aldridge, A., Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Bowler, L. (2004). Using human figure drawings to elicit information from alleged victims of child sexual abuse. *Journal of Consulting and Clinical Psychology, 72,* 304–316.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Prosecutors Research Institute. (2003). *Finding words: Half a nation by 2010. Interviewing children and preparing for court.* Retrieved from *www.ndaa.org/pdf/finding_words_2003/pdf.*

American Psychological Association. (1991). Minutes of the Council of Representatives. *American Psychologist, 46*, 722.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Apostal, R., & Marks, C. (1990). Correlations between the Strong–Campbell and Myers–Briggs scales of introversion–extraversion and career interests. *Psychological Reports, 66,* 811–816.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87,* 84–94.

Atkinson, L. (1986). The comparative validities of the Rorschach and MMPI: A meta-analysis. *Canadian Psychology, 27,* 238–247.

Ayearst, L. E., & Bagby, R. M. (2010). Evaluating the psychometric properties of psychological measures. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed., pp. 23–61). New York: Guilford Press.

Ball, J. D., Archer, R. P., & Imhof, E. A. (1994). Time requirements of psychological testing: A survey of practitioners. *Journal of Personality Assessment, 63,* 239–249.

Barbuto, J. E. (1997). A critique of the Myers–Briggs Type Indicator and its operationalization of Carl Jung's psychological types. *Psychological Reports, 80,* 611–625.

Boyle, G. J. (1995). Myers–Briggs Type Indicator (MBTI): Some psychometric limitations. *Australian Psychologist, 30,* 71–74.

Briggs, K. C., & Myers, I. B. (1998). *Myers–Briggs Type Indicator.* Menlo Park, CA: Consulting Psychologists Press.

Bruck, C. (2009). Human figure drawings and children's recall of touching. *Journal of Experimental Psychology: Applied, 15,* 361–374.

Buck, J. N. (1948). The HTP test. *Journal of Clinical Psychology, 4,* 151–159.

Burkitt, E., & Barrett, M. (2011). The effects of different drawing materials on children's drawings of positive and negative human figures. *Educational Psychology, 31,* 459–479.

Burns, R. C., & Kaufman, S. H. (1970). *Kinetic Family Drawings (K-F-D): An introduction to understanding children through kinetic drawings*. New York: Brunner/Mazel.

Butcher, J., & Rouse, S. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology, 47,* 87–111.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31,* 141–154.

Carlson, J. G. (1985). Recent assessments of the Myers–Briggs Type Indicator. *Journal of Personality Assessment, 49,* 356–365.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony.* Washington, DC: American Psychological Association.

Childs, R. A., & Eyde, L. D. (2002). Assessment training in clinical psychology doctoral programs: What should we teach? What do we teach? *Journal of Personality Assessment, 78,* 130–144.

Clemence, A. J., & Handler, L. (2001). Psychological assessment on internship: A survey of training directors and their expectations for students. *Journal of Personality Assessment, 76,* 18–47.

Conte, J. R., Sorenson, E., Fogarty, L., & Rosa, J. D. (1991). Evaluating children's reports of sexual abuse: Results from a survey of professionals. *American Journal of Orthopsychiatry, 61,* 428–437.

Cramer, P. (1991). *The development of defense mechanisms: Theory, research, and assessment.* New York: Springer-Verlag.

Cramer, P. (1999). Future directions for the Thematic Apperception Test. *Journal of Personality Assessment, 72,* 74–92.

Cramer, P. (2009). The development of defense mechanisms from pre-adolescence to early adulthood: Do IQ and social class matter? A longitudinal study. *Journal of Research in Personality, 43,* 464–471.

Cross, T. P., Jones, L. M., Walsh, W. A., Simone, M., & Kolko, D. (29007). Child forensic interviewing in Children's Advocacy Centers: Empirical data on a practice model. *Child Abuse and Neglect, 31,* 1031–1052.

Dana, R. H. (1985). Thematic Apperception Test (TAT). In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 89–134). Boston: Allyn & Bacon.

Dao, T. K., Prevatt, F., & Horne, H. L. (2008). Differentiating psychotic patients from nonpsychotic patients with the MMPI-2 and Rorschach. *Journal of Personality Assessment, 90,* 93–101.

Deaver, S. P. (2009). A normative study of children's drawings: Preliminary research

findings. *Art Therapy: Journal of the American Art Therapy Association, 26,* 4–11.

Dickinson, J. J., Poole, D. A., & Bruck, M. (2005). Back to the future: A comment on the use of anatomical dolls in forensic interviews. *Journal of Forensic Psychology Practice, 5,* 63–74.

Eron, L. D. (1950). A normative study of the Thematic Apperception Test. *Psychological Monographs, 64*(9), i–48.

Eron, L. D. (1953). Response of women to the Thematic Apperception Test. *Journal of Consulting Psychology, 17,* 269–282.

Everson, M. D., & Boat, B. W. (1994). Putting the anatomical doll controversy in perspective: An examination of the major uses and criticisms of the dolls in child sexual abuse evaluations. *Child Abuse and Neglect, 18,* 113–129.

Exner, J. E. (1969). *The Rorschach systems.* New York: Grune & Stratton.

Exner, J. E. (1974). *The Rorschach: A comprehensive system* (Vol. 1). New York: Wiley.

Exner, J. E. (1978). *The Rorschach: A Comprehensive System. Vol. 2. Current research and advanced interpretation.* New York: Wiley.

Exner, J. E. (1986). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (2nd ed.). New York: Wiley.

Exner, J. E. (1993). *The Rorschach: A comprehensive system. Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.

Exner, J. E. (1997). The future of the Rorschach in personality assessment. *Journal of Personality Assessment, 68,* 37–46.

Exner, J. E. (2001a). A comment on "The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach." *Clinical Psychology: Science and Practice, 8,* 386–388.

Exner, J. E. (2001b). *A Rorschach workbook for the Comprehensive System* (5th ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E. (2003). *The Rorschach: A Comprehensive System: Vol. 1. Basic foundations and principles of interpretation* (4th ed.). New York: Wiley.

Exner, J. E. (2007). A new U.S. adult nonpatient sample. *Journal of Personality Assessment, 89*(S1), S154–S158.

Exner, J. E., & Erdberg, P. (2005). *The Rorschach: A Comprehensive System: Vol. 2. Advanced interpretation* (3rd ed.). New York: Wiley.

Exner, J. E., & Exner, D. E. (1972). How clinicians use the Rorschach. *Journal of Personality Assessment, 36,* 403–408.

Faller, K. C. (2005). Anatomical dolls: Their use in assessment of children who may have been sexually abused. *Journal of Child Sexual Abuse, 14,* 1–21.

Feldman, M., & Hunt, R. G. (1958). A relation of difficulty in drawing and ratings of adjustment based on human figure drawings. *Journal of Consulting Psychology, 22,* 217–220.

Flanagan, R., & Motta, R. W. (2007). Figure drawings: A popular method. *Psychology in the Schools, 44,* 257–270.

Fleiss, J. (1981). *Statistical methods for rates and proportions* (2nd ed.). Hoboken, NJ: Wiley.

Freedenfeld, R. N., Ornduff, S. R., & Kelsey, R. M. (1995). Object relations and physical abuse: A TAT analysis. *Journal of Personality Assessment, 64,* 552–568.

Furnham, A. (1996). The Big Five versus the Big Four: The relationship between

the Myers–Briggs Type Indicator (MBTI) and the NEO-PI five factor model of personality. *Personality and Individual Differences, 21,* 303–307.
Furnham, A., & Crump, J. (2007). Relationship between the MBTI and FIRO-B in a large British sample. *Psychological Reports, 101*, 970–978.
Furnham, A., Dissou, G., Sloan, P., & Chamorro-Premuzic, T. (2007). Personality and intelligence in business people: A study of two personality and two intelligence measures. *Journal of Business and Psychology, 22*, 99–109.
Furnham, A., & Stringfield, P. (1993). Personality and work performance: Myers–Briggs Type Indicator correlates of managerial performance in two cultures. *Personality and Individual Differences, 14,* 145–153.
Ganellen, R. J. (1996). Exploring MMPI–Rorschach relationships. *Journal of Personality Assessment, 67,* 529–542.
Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review, 4,* 641–655.
Garb, H. N. (1998). Recommendations for training in the use of the Thematic Apperception Test (TAT). *Professional Psychology: Research and Practice, 29,* 621–622.
Garb, H. N. (1999). Call for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings. *Assessment, 6,* 313–317.
Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from meta-analyses. *Psychological Science, 9,* 402–404.
Garb, H. N., Lilienfeld, S., & Fowler, K. A. (2008). Psychological assessment and clinical judgment. In J. E. Maddux & B. A. Winstead (Eds.), *Psychopathology: Foundations for a contemporary understanding* (2nd ed., pp. 103–124). New York: Routledge.
Garb, H. N., Wood, J. M., & Nezworski, M. T. (2000). Projective techniques and the detection of child sexual abuse. *Child Maltreatment, 5,* 161–168.
Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Towards a resolution of the Rorschach controversy. *Psychological Assessment, 13*, 433–448.
Garden, A.-M. (1991). Unresolved issues with the Myers–Briggs Type Indicator. *Journal of Psychological Type, 22,* 3–14.
Girelli, S. A., & Stake, J. E. (1993). Bipolarity in Jungian type theory and the Myers–Briggs Type Indicator. *Journal of Personality Assessment, 60,* 290–301.
Gronnerod, C. (2003). Temporal stability in the Rorschach method: A meta-analytic review. *Journal of Personality Assessment, 80,* 272–293.
Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). New York: Wiley.
Hamel, M., Shaffer, T. W., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment, 75,* 280–294.
Hammer, E. F. (1985). The House–Tree–Person test. In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 135–164). Boston: Allyn & Bacon.
Handler, L. (1985). The clinical use of the Draw-A-Person test (DAP). In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 165–216). Boston: Allyn & Bacon.
Handler, L., & Habenicht, D. (1994). The Kinetic Family Drawing: A review of the literature. *Journal of Personality Assessment, 62,* 440–464.

Harris, D. B. (1963). *Children's drawings as a measure of intellectual maturity.* New York: Harcourt, Brace & Wood.

Harvey, R. J., Murry, W. D., & Stamoulis, D. T. (1995). Unresolved issues in the dimensionality of the Myers–Briggs Type Indicator. *Educational and Psychological Measurement, 55,* 535–544.

Haynes, S. N., Smith, G., & Hunsley, J. (2011). *Scientific foundations of clinical assessment.* New York: Taylor & Francis.

Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11,* 278–296.

Hilsenroth, M. J., & Handler, L. (1995). A survey of graduate students' experiences, interests, and attitudes about learning the Rorschach. *Journal of Personality Assessment, 64,* 243–257.

Hlavka, H. R., Olinger, S. D., & Lashley, J. L. (2010). The use of anatomical dolls as a demonstration aid in child sexual abuse interviews: A study of forensic interviewers' perceptions. *Journal of Child Sexual Abuse, 19,* 519–553.

Holt, R. R. (1999). Empiricism and the Thematic Apperception Test: Validity is the payoff. In L. Gieser & M. I. Stein (Eds.), *Evocative images: The Thematic Apperception Test and the art of projection* (pp. 99–105). Washington, DC: American Psychological Association.

Holtzman, W. H. (1993). An unjustified, sweeping indictment by Motta et al. of human figure drawings for assessing psychological functioning. *School Psychology Quarterly, 8,* 189–190.

Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment, 11,* 266–277.

Hunsley, J., & Bailey, J. M. (2001). Whither the Rorschach?: An analysis of the evidence. *Psychological Assessment, 13,* 472–485.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3,* 29–51.

Hunsley, J., & Mash, E. J. (2008). Developing criteria for evidence-based assessment: An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3–14). New York: Oxford University Press.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15,* 446–455.

Jackson, S. L., Parker, C. P., & Dipboye, R. L. (1996). A comparison of competing models underlying responses to the Myers–Briggs Type Indicator. *Journal of Career Assessment, 4,* 99–115.

Jenkins, S. R. (Ed.). (2008). *A handbook of clinical scoring systems for thematic apperceptive techniques.* Mahwah, NJ: Erlbaum.

Joiner, T. E., Schmidt, K. L., & Barnett, J. (1996). Size, detail, and line heaviness in children's drawings as correlates of emotional distress: (More) negative evidence. *Journal of Personality Assessment, 67,* 127–141.

Jørgensen, K., Andersen, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach depression index and the schizophrenia index: A review. *Assessment, 7,* 259–280.

Jørgensen, K., Andersen, T. J., & Dam, H. (2001). "The diagnostic efficiency of the Rorschach depression index and the schizophrenia index: A review": Erratum. *Assessment, 8,* 355.

Katz, H. E., Russ, S. W., & Overholser, J. C. (1993). Sex differences, sex roles, and projection on the TAT: Matching stimulus to examinee gender. *Journal of Personality Assessment, 60,* 186–191.

Kaufman, A. S., Kaufman, N. L., & McLean, J. E. (1993). Profiles of Hispanic adolescents and adults on the Myers–Briggs Type Indicator. *Perceptual and Motor Skills, 76,* 628–630.

Keiser, R. E., & Prather, E. N. (1990). What is the TAT?: A review of ten years of research. *Journal of Personality Assessment, 55,* 800–803.

Koocher, G. P., Goodman, G. S., White, C. S., Friedrich, W. N., Sivan, A. B., & Reynolds, C. R. (1995). Psychological science and the use of anatomically detailed dolls in child sexual abuse assessments. *Psychological Bulletin, 118,* 199–122.

Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings.* New York: Grune & Stratton.

Lally, S. J. (2001). Should human figure drawings be admitted into court? *Journal of Personality Assessment, 76*, 135–149.

Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations?: A survey of experts. *Professional Psychology: Research and Practice, 34,* 491–498.

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48,* 72–79.

Lett, J. (1990). A field guide to critical thinking. *Skeptical Inquirer, 14*(2), 153–160.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1,* 27–66.

Lis, A., Parolin, L., Calvo, V., Zennaro, A., & Meyer, G. (2007). The impact of administration and inquiry on Rorschach Comprehensive System protocols in a national reference sample. *Journal of Personality Assessment, 89,* S193–S200.

Lorr, M. (1991). An empirical evaluation of the MBTI typology. *Personality and Individual Differences, 12,* 1141–1145.

MacDonald, D. A., Anderson, P. E., Tsagarakis, C. I., & Holland, C. J. (1994). Examination of the relationship between the Myers–Briggs Type Indicator and the NEO Personality Inventory. *Psychological Reports, 74,* 339–344.

Machover, K. (1949). *Personality projection in the drawing of the human figure.* Springfield, IL: Charles C Thomas.

Masling, J. M. (1992). The influence of situation and interpersonal variables in projective testing. *Journal of Personality Assessment, 59,* 616–640.

Masling, J. M. (1997). On the nature and utility of projective tests and objective tests. *Journal of Personality Assessment, 69,* 257–270.

Matthews, P. R. (2004). Response to Clack and Allen. Retrieved June 13, 2012, from *www.bmj.com/rapid-response/2011/10/30/response-clack-amp-allen-4*.

Matto, H. C. (2002). Investigating the validity of the Draw-a-Person: Screening Procedure for Emotional Disturbance: A measurement validation study with high-risk youth. *Psychological Assessment, 14*(2), 221–225.

Matto, H. C., & Naglieri, J. A. (2005). Race and ethnic differences and human figure drawings: Clinical utility of the DAP:SPED. *Journal of Clinical Child and Adolescent Psychology, 34,* 706–711.

McCaulley, M. H., & Martin, C. R. (1995). Career assessment and the Myers–Briggs Type Indicator. *Journal of Career Assessment, 3,* 219–239.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Bulletin, 96,* 690–702.

McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers–Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57,* 17–40.

McFall, R. M. (2005). Theory and utility—key themes in evidence-based assessment: Comment on the special section. *Psychological Assessment, 17,* 312–323.

McGrath, R. E., & Carroll, E. J. (2012). The current status of "projective" "tests." In H. Cooper (Ed.), *APA handbook of research methods in psychology: Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 329–348). Washington, DC: American Psychological Association.

McGrath, R. E., Pogge, D. L., Stokes, J. M., Cragnolino, A., Zaccario, M., Hayman, J., et al. (2005). Field reliability of Comprehensive System scoring in an adolescent inpatient sample. *Assessment, 12*, 199–209.

Meyer, G. J. (1991). An empirical search for fundamental personality and mood dimensions within the Rorschach test (Doctoral dissertation, Loyola University of Chicago, 1991). *Dissertation Abstracts International, 52,* 1071B–1072B.

Meyer, G. J. (1997). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment, 9,* 480–489.

Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment, 13,* 486–502.

Meyer, G. J., Erdberg, P., & Shaffer, T. W. (2007). Toward international normative reference data for the Comprehensive System. *Journal of Personality Assessment, 89,* S201–S216.

Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Fowler, J. C., Piers, C. C., et al. (2002). An examination of interrater reliability for scoring the Rorschach Comprehensive System in eight data sets. *Journal of Personality Assessment, 78*, 219–274.

Meyer, G. J., & Viglione, D. J. (2008). An introduction to Rorschach assessment. In R. P. Archer & S. R. Smith (Eds.), *A guide to personality assessment: Evaluation, application, and integration* (pp. 281–336). New York: Routledge.

Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation, and technical manual*. Toledo, OH: Rorschach Performance Assessment System, LLC.

Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive System. *Psychological Bulletin*, *139*, 548–605

Motta, R. W., Little, S. G., & Tobin, M. I. (1993). The use and abuse of human figure drawings. *School Psychology Quarterly, 8,* 162–169.

Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.

Murstein, B. L. (1965). *Handbook of projective techniques*. New York: Basic Books.

Murstein, B. L. (1972). Normative written TAT responses for a college sample. *Journal of Personality Assessment, 36,* 104–107.

Naglieri, J. A., & Pfeiffer, S. I. (1992). Performance of disruptive behavior-disordered and normal samples on the Draw A Person: Screening Procedure for Emotional Disturbance. *Psychological Assessment, 4,* 156–159.

National District Attorneys Association. (2010). *Legislation permitting the use of anatomical dolls in child abuse cases* (Updated June 2010). Retrieved from *www.ndaa.org/ncpa.state_statutes.html*.

Nichols, R. C., & Strumpfer, D. J. (1962). A factor analysis of Draw-A-Person test scores. *Journal of Consulting Psychology, 26,* 156–161.

Nunnally, J. C., & Bernstein, I. C. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Palmer, L., Farrar, A. R., Valle, M., Ghahary, N., Panella, M., & DeGraw, D. (2000). An investigation of the clinical use of the House–Tree–Person projective drawings in the psychological evaluation of child sexual abuse. *Child Maltreatment, 5,* 169–175.

Parker, K. C. H. (1983). A meta-analysis of the reliability and validity of the Rorschach. *Journal of Personality Assessment, 47,* 227–231.

Parker, K. C. H., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin, 103,* 367–373.

Paul, A. M. (2004). *The cult of personality: How personality tests are leading us to miseducate our children, mismanage our companies, and misunderstand ourselves*. New York: Free Press.

Pearman, R. R., & Fleenor, J. (1996). Differences in observed and self-reported qualities of psychological types. *Journal of Psychological Type, 39,* 3–17.

Peterson, C., & Ulrey, L. M. (1994). Can explanatory style be scored from TAT protocols? *Personality and Individual Differences, 20,* 102–106.

Pinkerman, J. E., Haynes, J. P, & Keiser, T. (1993). Characteristics of psychological practice in juvenile court clinics. *American Journal of Forensic Psychology, 11*(2), 3–12.

Pipe, M.-E., & Salmon, K. (2009). Dolls, drawings, body diagrams, and other props: Role of props in investigative interviews. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 365–395). Hoboken, NJ: Wiley.

Pittenger, D. J. (1993). The utility of the Myers–Briggs Type Indicator. *Review of Educational Research, 63,* 467–488.

Pittenger, D. J. (2005). Cautionary comments regarding the Myers–Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research, 57,* 210–221.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Psychological Corporation. (1997). *Wechsler Adult Intelligence Scale, Third Edition: Technical manual*. San Antonio, TX: Author.

Rapaport, D., Gill, M., & Shafer, R. (1946). *Diagnostic psychological testing*. Chicago: Year Book.

Reznikoff, M., Aronow, E., & Rauchway, A. (1982). The reliability of inkblot content scales. In C. D. Spielberger & J. D. Butcher (Eds.), *Advances in personality assessment* (Vol. 1, pp. 83–113). Hillsdale, NJ: Erlbaum.

Riethmiller, R. J., & Handler, L. (1997). Problematic methods and unwarranted conclusions in DAP research: Suggestions for improved research procedures. *Journal of Personality Assessment, 69,* 459–475.

Riordan, R., & Verdel, A. C. (1991). Evidence of sexual abuse in children's art products. *School Counselor, 39,* 116–121.

Ritsher, J. B. (2004). Association of Rorschach and MMPI psychosis indicators

and schizophrenia spectrum diagnoses in a Russian clinical sample. *Journal of Personality Assessment, 83,* 46–63.

Ronan, G. F., Gibbs, M. S., Dreer, L. E, & Lombardo, J. A. (2008). Personal problem-solving system-revised. In S. R. Jenkins (Ed.), *A handbook of clinical scoring systems for thematic apperceptive techniques* (pp. 181–207). Mahwah, NJ: Erlbaum.

Rorschach, H. (1921/1964). *Psychodiagnostics*. New York: Grune & Stratton. (Original work published in German in 1921 and in English in 1942)

Rossini, E. D., & Moretti, R. J. (1997). Thematic Apperception Test (TAT) interpretation: Practice recommendations from a survey of clinical psychology doctoral programs accredited by the American Psychological Association. *Professional Psychology: Research and Practice, 28,* 393–398.

Sattler, J. M. (1998). *Clinical and forensic interviewing of children and families: Guidelines for the mental health, education, pediatric, and child maltreatment fields*. San Diego, CA: Author.

Schaubhut, N. A., Herk, N. A., & Thompson, R. C. (2009). *MBTI Form M manual supplement*. Menlo Park, CA: Consulting Psychologists Press.

Schaubhut, N. A., & Thompson, R. C. (2011). *MBTI Step II manual supplement*. Menlo Park, CA: Consulting Psychologists Press.

Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment, 73,* 305–316.

Shontz, F. C., & Green, P. (1992). Trends in research on the Rorschach: Review and conclusions. *Applied and Preventive Psychology, 1,* 149–156.

Sipps, G. J., Alexander, R. A., & Friedt, L. (1985). Item analysis of the Myers–Briggs Type Indicator. *Educational and Psychological Measurement, 45,* 789–796.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112,* 140–154.

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology, 5,* 89–113.

Sultan, S., Andronikof, A., Réveillère, C., & Lemmel, G. (2006). A Rorschach stability study in a nonpatient adult sample. *Journal of Personality Assessment, 87,* 113–119.

Sultan, S. & Meyer, G. J. (2009). Does productivity impact the stability of Rorschach scores? *Journal of Personality Assessment, 91*, 480–493.

Teglasi, H. (2010). *Essentials of TAT and other storytelling assessments* (2nd ed.). Hoboken, NJ: Wiley.

ter Laak, J., de Goede, M., Aleva, A., & van Rijswijk, P. (2005). The Draw-A-Person test: An indicator of children's cognitive and socioemotional adapatation? *Journal of Genetic Psychology, 66,* 77–93.

Tharinger, D. J., & Stark, K. (1990). A qualitative versus quantitative approach to evaluating the Draw-A-Person and Kinetic Family Drawing: A study of mood and anxiety-disorder children. *Psychological Assessment, 2,* 365–375.

Thierry, K. L., Lamb, M. E., Orbach, Y., & Pipe, M.-E. (2005). Developmental differences in the function and use of anatomical dolls during interviews with alleged sexual abuse victims. *Journal of Consulting and Clinical Psychology, 73,* 1125–1134.

Thomas, G. V., & Jolley, R. P. (1998). Drawing conclusions: A re-examination of

empirical and conceptual bases for psychological evaluations of children from their drawings. *British Journal of Clinical Psychology, 37,* 127–139.

Thompson, B., & Borrello, G. M. (1986). Second-order factor structure of the MBTI: A construct validity assessment. *Measurement and Evaluation in Counseling and Development, 18,* 148–153.

Tischler, L. (1994). The MBTI factor structure. *Journal of Psychological Type, 31,* 24–31.

Turcotte, M. (1994). Use of tests in the Canadian employment services. *Journal of Employment Counseling, 31,* 188–199.

Tzeng, O. C., Ware, R., & Chen, J.-M. (1989). Measurement and utility of continuous unipolar ratings for the Myers–Briggs Type Indicator. *Journal of Personality Assessment, 53,* 727–738.

Vacha-Haase, T., Henson, R., & Caruso, J. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, *62*, 562–569.

Vane, J. R. (1981). The Thematic Apperception Test: A review. *Clinical Psychology Review, 1,* 319–336.

Vass, Z. (1998). The inner formal structure of the H-T-P Drawings: An exploratory study. *Journal of Clinical Psychology, 54,* 611–619.

Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment, 11,* 251–263.

Viglione, D. J., & Hilsenroth, M. J. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment, 13,* 452–471.

Viglione, D. J., & Meyer, G. J. (2008). An overview of Rorschach psychometrics for forensic practice. In C. B. Gacono, F. B. Evans, N. Kaser-Boyd, & L. A. Gacono (Eds.), *The handbook of forensic Rorschach assessment* (pp. 21–53). New York: Routledge.

Viglione, D. J., & Taylor, N. (2003). Empirical support for interrater reliability of Rorschach Comprehensive System coding. *Journal of Clinical Psychology, 59,* 111–121.

Vitacco, M. J., Lilienfeld, S. O., Erickson, S., & Wood, J. M. (2012). Challenging personality testing: Objective and projective instruments. In D. Faust (Ed.), *Coping with psychiatric and psychological testimony* (6th ed., pp. 335–362). New York: Oxford University Press.

Wanderer, Z. W. (1997). Validity of clinical judgments based on human figure drawings. In E. F. Hammer (Ed.), *Advances in projective drawing interpretation* (pp. 301–315). Springfield, IL: Charles C Thomas.

Watkins, C. E., Jr., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26,* 54–60.

Weiner, I. B. (1966). *Psychodiagnosis in schizophrenia*. New York: Wiley.

Weiner, I. B. (1993). Clinical considerations in the conjoint use of the Rorschach and the MMPI. *Journal of Personality Assessment, 60,* 148–152.

Weiner, I. B. (1994). The Rorschach Inkblot Method (RIM) is not a test: Implications for theory and practice. *Journal of Personality Assessment, 62,* 498–504.

Weiner, I. B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment, 68,* 5–19.

Weiner, I. B. (1999). What the Rorschach can do for you: Incremental validity in clinical applications. *Assessment, 6,* 327–339.

Weiner, I. B. (2001). Considerations in collecting Rorschach reference data. *Journal of Personality Assessment, 77,* 122–127.

West, M. M. (1998). Meta-analysis of studies assessing the efficacy of projective techniques in discriminating child sexual abuse. *Child Abuse and Neglect, 11,* 1151–1166.

Westen, D. (1991). Clinical assessment of object relations using the TAT. *Journal of Personality Assessment, 56,* 56–74.

Westen, D., Lohr, N. E., Silk, K., Gold, L., & Kerber, K. (1990). Object relations and social cognition in borderlines, major depressives, and normals: A Thematic Apperception Test analysis. *Psychological Assessment: Journal of Consulting and Clinical Psychology, 2,* 355–364.

Widiger, T. A., & Schilling, K. M. (1980). Toward a construct validation of the Rorschach. *Journal of Personality Assessment, 44,* 450–459.

Williams, S. D., Wiener, J., & MacMillan, H. (2005). Build-a-person technique: An examination of the validity of human-figure features as evidence of childhood sexual abuse. *Child Abuse and Neglect, 29,* 701–713.

Wolfner, G., Faust, D., & Dawes, R. M. (1993). The use of anatomically detailed dolls in sexual abuse evaluations: The state of the science. *Applied and Preventive Psychology, 2,* 1–11.

Wood, J. M., Garb, H. N., & Nezworski, M. T. (2006). Psychometrics: Better measurement makes better clinicians. In S. O. Lilienfeld & W. T. O'Donohue (Eds.), *The great ideas of clinical science: The 17 concepts that every mental health practitioner should understand* (pp. 77–92). New York: Brunner-Routledge.

Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000). The Rorschach test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology, 56,* 395–430.

Wood, J. M., Lilienfeld, S. O., Nezworski, M. T., Garb, H. N., Allen, K. H., & Wildermuth, J. (2010). Validity of Rorschach Inkblot scores for discriminating psychopaths from nonpsychopaths in forensic populations: A meta-analysis. *Psychological Assessment, 22,* 336–349.

Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001). The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach. *Clinical Psychology: Science and Practice, 8,* 350–373.

Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2006, Spring). The controversy over Exner's comprehensive system for the Rorschach: The critics speak. *The Independent Practitioner, 26*(2), 73–82.

Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2003). *What's wrong with the Rorschach?: Science confronts the controversial inkblot test.* San Francisco: Jossey-Bass.

Wood, J. A., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2009). Projective techniques in the courtroom. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Controversies and consensus* (pp. 202–223). New York: Guilford Press.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The comprehensive system for the Rorschach: A critical examination. *Psychological Science, 7,* 3–10.

CHAPTER FOUR

The Science and Pseudoscience of Expert Testimony

Joseph T. McCann, Steven Jay Lynn,
Scott O. Lilienfeld, Kelley L. Shindler,
and Tammy R. Hammond Natof

Over the past several decades mental health professionals, including clinical psychologists, have played an increasingly prominent role as expert witnesses in legal proceedings. Several factors account for this trend, including more mental health professionals turning to forensic work as a way of coping with managed care and greater attention being paid by the courts to serious social problems, such as child abuse, sexual harassment, domestic violence, and the prediction of violence risk among psychiatric patients.

According to Ciccone (1992), the use of expert witnesses dates back many centuries to early Egyptian and Greek societies that used forensic medicine to resolve legal issues. Over time, litigants and their attorneys have acquired greater control over the selection, preparation, and presentation of expert witnesses in the courts (Landsman, 1995). The use of expert testimony pertaining to issues involving psychological or mental health issues also has a lengthy history. Cases involving questions of a criminal defendant's sanity, for example, have been at the forefront of many debates about the appropriate role of mental health professionals in legal proceedings. In 1843, Daniel McNaughton shot and killed the secretary to the British prime minister and was found not guilty by reason of insanity (Steadman et al., 1993). There was considerable public outcry over the verdict, although the legal test for insanity arising from this case has endured.

The use of behavioral science and mental health expert testimony has expanded considerably over the last few decades to include such issues as competency to stand trial, diminished capacity, mitigation and aggravating factors at sentencing, violence risk and personality assessment, and the accuracy of eyewitness testimony. In civil and family court cases, expert testimony is frequently offered on issues such as claims of psychological injury or neuropsychological impairment, the veracity of child sexual abuse allegations, and domestic violence. Along with the expanding use of experts in court has come considerable debate about the reliability, validity, and appropriateness of permitting experts from the behavioral sciences to testify in court proceedings, as well as the scientific basis of theories, constructs, and diagnostic syndromes that form the basis of their testimony (Faigman, Saks, Sanders, & Cheng, 2008; Faust, 2012). Among the many challenges confronting expert witnesses is that jurors tend to be more influenced by expert testimony based on clinical opinions than on actuarial data (Krauss & Sales, 2001). Accordingly, it is crucial that experts be persuasive and credible to triers of fact when presenting research-based evidence.

In this chapter we review legal and professional standards pertaining to the admissibility of expert testimony by behavioral scientists, including mental health professionals. More specifically, we review the extant standards for admissibility of evidence in state and federal courts, as well as professional guidelines that govern the ethics and standards of such testimony. We then review several major areas addressed by expert testimony, including the accuracy of eyewitnesses, psychiatric diagnoses and syndromes, psychometric testing, and the prediction of violent behavior. In addition, we review several controversial syndromes and constructs that have either been proposed or proffered in legal settings. Through this discussion, we attempt to provide some guidance for distinguishing between those areas that have an adequate scientific basis from those that are pseudoscientific and have questionable validity.

ADMISSIBILITY OF SCIENTIFIC EVIDENCE IN THE COURTROOM

Legal Standards

The determination of whether evidence is admitted in any legal proceeding is governed by rules of evidence. Because the legal system in the United States is federalist in nature, in that there are separate state and federal courts, evidentiary rules can vary across different jurisdictions. Federal courts rely on the *Federal Rules of Evidence* (FRE; Green & Nesson, 1992), whereas state courts typically rely on either codified rules of evidence, which in some instances may be modeled after the FRE, or an extensive body of case law. Regardless of whether the specific jurisdiction is state or federal, one of two major legal standards typically governs the admissibility of expert testimony. One standard is the *Frye* test, which was originally outlined in *United*

States v. Frye (1923) and the other is the FRE, which was further expanded upon in the landmark case *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993), also known as the *Daubert* standard. Because of its importance, as well as some inconsistencies in the manner in which principles outlined in *Daubert* were applied by various courts, several important post-*Daubert* decisions set forth by the United States Supreme Court have clarified the standards for admissibility of expert testimony in federal courts.

In this section, we outline the major evidentiary standards, including the *Frye* test and the FRE. In addition, we provide a brief review of the standards outlined in the Supreme Court's *Daubert* decision as well as an overview of post-*Daubert* opinions. Finally, we briefly review the status of each standard across various state and federal jurisdictions. Although statutes and case law are subject to change and the material presented herein is not intended to serve as legal advice, at least one of these general standards on the admissibility of expert testimony is likely to be applicable in any given jurisdiction in the United States.

The *Frye* Test

One of the major legal standards for the admissibility of expert testimony was established in the case of *United States v. Frye* (1923). Although the *Frye* decision was a federal opinion issued in the early part of the 20th century by the Court of Appeals of the District of Columbia, the *Frye* test was widely adopted by other federal jurisdictions and many state courts. Interestingly, the *Frye* test, also referred to as the "general acceptance" standard, arose from a case addressing the admissibility of a systolic blood pressure deception test, which was a precursor to the polygraph or so-called lie detector test. In ruling that this test was not admissible because it lacked adequate scientific recognition, the court in *Frye* stated:

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable states is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained *general acceptance in the particular field in which it belongs*. (*United States v. Frye*, 1923, p. 1014, emphasis added)

When adopting the *Frye* standard, courts typically examine whether a particular theory, technique, diagnostic classification, or methodology is generally accepted within the specific field from which the expert's testimony is derived. Some examples of how general acceptance can be established include examining peer-reviewed literature, surveys of common practices used by professionals, and scholarly treatises or books.

Federal Rules of Evidence

In 1961, the Chief Justice of the United States Supreme Court, Earl Warren, appointed a special committee to study whether a formal set of federal evidentiary laws was both possible and desirable (Green & Nesson, 1992). The FRE were signed into law in 1975, and they have become the formal legal standard for determining the admissibility of evidence in federal courts. Many states have enacted codes of evidence that are largely based on the FRE, making them "the most significant single source of evidence law in America" (Green & Nesson, 1992, p. xii).

The admissibility of expert testimony is governed by Article VII of the FRE, which includes Rule 702:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

Therefore, the FRE can be conceptualized as a "helpfulness" standard in that expert testimony which is of assistance to the trier of fact (i.e., judge or jury) in understanding evidence or factual circumstances in a case is admissible.

Rule 703 of the FRE pertains to the basis of an expert witness-opinion:

> The facts or data in the particular case on which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the participant, the facts or data need not be admissible in evidence.

Therefore, when judging the admissibility of an expert's opinion, the court can examine the reasonableness of the expert's reliance on a particular technique, theory, and body of research by examining the practices of other experts in the field.

Two additional provisions in the FRE pertaining to expert testimony are worth noting. Rule 704 governs the controversial issue of whether experts may offer opinions on the "ultimate issue," such as whether a specific accident caused the plaintiff's injuries, or whether a criminal defendant is competent to stand trial. Rule 704 of the FRE states that experts can testify as to the ultimate issue, with the exception that they may not offer an opinion regarding whether the defendant's mental state or condition played a direct role in some aspect of the crime (Green & Nesson, 1992). This rule essentially overrules previous case law that limited the extent to which experts could testify on ultimate issues (Goodman-Delahunty, 1997). According to Green and Nesson (1992), however, Rule 704 was drafted to facilitate the admissibility of ultimate issue testimony because

restricting such testimony "generally served to deprive the trier of fact of useful information" (p. 132). The sole restriction of ultimate opinion testimony under Rule 704 pertains to the issue of a federal criminal defendant's mental state during commission of an offense and was implemented as part of the Insanity Defense Reform Act of 1984 that followed the controversial acquittal of John Hinckley Jr. for the assassination attempt on President Ronald Reagan (McPherson, 1999; Steadman et al., 1993). In addition, Rule 705 of the FRE pertains to the disclosure of facts or data on which the expert relied in forming his or her opinion. In this context, facts or data would refer to research studies, data from clinical examinations, and other information that the expert utilized when forming an opinion. According to Rule 705, experts do not have to disclose such facts or data before offering their opinions, unless the court requires them to do so, although they may be required to disclose them on cross-examination.

The *Daubert* Standard

A third major standard for determining the admissibility of expert testimony was outlined by the United States Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993). The *Daubert* opinion arose from a civil case in which a pharmaceutical company was sued for damages related to birth defects that were allegedly caused by pregnant mothers ingesting an antinausea drug manufactured by the company. The major focus of the *Daubert* opinion was the appropriate standard for admissibility of expert testimony offered by the plaintiffs; the testimony was based on scientific principles determined by the trial court that failed to meet the "general acceptance" test outlined in *United States v. Frye* (1923).

Both the FRE and *Frye* test are federal standards governing admissibility, one being a statutory code (i.e., FRE) and the other a published legal opinion (i.e., *Frye*). The issue in *Daubert* was a determination of the appropriate standard for admissibility in federal courts, given that federal jurisdictions had been essentially divided on whether the FRE or *Frye* test was the appropriate standard. The United States Supreme Court held that "the Federal Rules of Evidence, not *Frye*, provide the standard for admitting expert scientific testimony in a federal trial" (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993, p. 2790). Moreover, the *Daubert* opinion held that the FRE "assign[s] to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand" (p. 2799).

The most interesting aspect of the *Daubert* decision, and the one of greatest relevance to expert witnesses, is the delineation of four major factors that trial judges may consider when determining whether expert testimony is admissible, notably: (1) The theory or technique is scientific knowledge that is testable; (2) the theory or technique has been subjected to peer review; (3) the rate of potential error is known; and (4) the theory

or technique has gained general acceptance in the field. With respect to the first of these factors, the Supreme Court stated that the critical question is whether hypotheses can be developed and tested to evaluate their validity. The second factor recommends that judges determine the extent to which a theory or technique has been subjected to peer review. Peer review helps to "increase the likelihood that substantive flaws in methodology will be detected" (p. 2797). The third factor, error rate, applies to "particular" scientific techniques and refers also to the "existence and maintenance of standards controlling the technique's operation" (p. 2797). Fourth and finally, the general acceptance of a theory or technique, as outlined earlier in *Frye*, remains an important consideration in determining the admissibility of expert testimony, but under *Daubert* general acceptance is no longer the sole determining factor for admissibility.

Considerable academic discussion has focused on the implications of the *Daubert* decision (e.g., Goodman-Delahunty, 1997; Lubit, 1998). The FRE standard of helpfulness has typically been construed as a more liberal standard that permits greater leniency when deciding what evidence is admissible, whereas the *Frye* general acceptance test has been construed as more stringent (Blau, 1998). Given that the *Daubert* standard offers specific factors that judges may consider when deciding on the admissibility of expert testimony, some questions raised by the *Daubert* decision required further elaboration. One such issue was whether *Daubert* would make it more difficult for certain types of expert testimony to be ruled admissible, thus making the *Daubert* standard more stringent than the liberal FRE standard dictates. Evidence suggests that, in general, judges are more likely to exclude scientific evidence now than before *Daubert* (Faigman & Monahan, 2009; McAuliff & Groscup, 2009).

With respect to the scope of the *Daubert* standard, another issue was raised as to whether the social sciences, including psychology, would be characterized as "scientific, technical, or other knowledge" as outlined in Rule 702 of the FRE (Faigman, 1995). In addition, the *Daubert* decision placed judges, who are often untrained in science, in the gatekeeper role of determining the scientific adequacy of expert testimony. Nevertheless, in addition to jury-eligible community members (McAuliff, Kovera, & Nunez, 2009; Cutler & Kovera, 2011) and attorneys (Kovera & McAuliff, 2009), judges (Gatowski et al., 2001) exhibit quite limited and unsophisticated understanding of scientific concepts (reliability, falsifiability, control groups) essential to fulfilling their gatekeeper role. More recent case law has clarified some issues unresolved by *Daubert*.

Post-*Daubert* Rulings

In the U.S Supreme Court case of *General Electric Co. v. Joiner* (1997), the gatekeeper role of the trial judge outlined in *Daubert* was affirmed. More important, the *Joiner* decision held that an "abuse of discretion"

standard should be applied to appellate review of rulings of the trial judge on the admissibility of expert testimony. *Joiner* appears to hold that "trial courts have broad discretion to reject proffered expert opinions if they are inadequately supported by the data" (Grudzinskas & Appelbaum, 1998, p. 502). The *Joiner* opinion also underscored the ruling in *Daubert* that even though the FRE replaced *Frye*, this did not mean that no limits could be placed on the admissibility of expert testimony.

A second major Supreme Court case on the *Daubert* standard was outlined in *Kumho Tire Co., Ltd. v. Carmichael* (1999). The case involved civil litigation over personal injuries resulting from a traffic accident in which the plaintiffs offered expert testimony on alleged defects in a tire on one of the vehicles. The significant holding in *Kumho* was that the admissibility of the expert's testimony, even though characterized as "technical" and not necessarily scientific, was to be analyzed according to the *Daubert* criteria. As a result of this ruling, the *Daubert* standard applied to all expert testimony, not just to scientific testimony (Cavanagh, 1999). Moreover, the *Kuhmo* ruling reemphasized the gatekeeper role of the trial judge and the need to apply the *Daubert* criteria flexibly.

As a result of the rulings in *Joiner* and *Kuhmo*, trial judges in federal courts are empowered to exercise discretion when excluding junk science from the courtroom, while permitting expert testimony that is relevant, reliable, and based on acceptable methodologies (Cavanagh, 1999; Littleton, 1999). Grudzinskas (1999) noted that these recent Supreme Court cases included the following implications: (1) opinions should be based on data, (2) inferences should be differentiated from data, (3) considerations should be given to the selection of methodologies used to collect data, and (4) the expert should be prepared to support the appropriateness of the methodology. These factors extend to the selection of methodologies (e.g., direct interviews, collateral interviews, psychological testing), diagnostic concepts or terminology, and explanatory theories.

Standards for Admissibility across Jurisdictions

The two major standards that remain for judging the admissibility of expert testimony are the general acceptance test outlined in *Frye* and the relevant, helpful, and reliable standard outlined in the FRE and elaborated in *Daubert*. Although the FRE and *Frye* are both federal standards, most state courts have adopted one of these standards. Because state courts are not bound by the rulings of federal courts, there may be considerable variability in legal standards across states as well as between federal and state jurisdictions within the same state. Therefore, state courts vary in their adoption of the *Frye v*ersus *Daubert* standards. As of this writing, 32 states broadly follow the *Daubert* standard (see Ambrogi, 2011). Some states (e.g., New York) have adopted two different standards, with the *Frye* test applicable in state courts and the *Daubert* standard applicable in federal

courts within the state. Appellate courts retain "remarkable power to find expert testimony inadmissible as a matter of law and to direct a verdict accordingly" (Faigman & Monahan, 2009, p. 8). No legal standard can prevent flawed science from entering the courtroom (McAuliff & Groscup, 2009). Accordingly, experts able to distinguish science from pseudoscience play an invaluable role in serving the judicial process.

Professional Standards for Expert Testimony

Although legal standards provide the ultimate test for determining whether an expert's testimony is admitted, professional standards provide guidance as well. The Committee on Ethical Guidelines for Forensic Psychologists of Division 41 of the American Psychological Association developed *Specialty Guidelines for Forensic Psychology* (2013) to inform forensic practice. With respect to expert testimony, these guidelines echo many of the provisions of the American Psychological Association (2002) ethical code, such as avoiding dual relationships and providing adequate documentation for one's conclusions. Furthermore, the *Specialty Guidelines* recommend that forensic psychologists take steps to ensure that their work, including communications in the form of reports or testimony, is not misrepresented or misused in legal proceedings. Testimony must be provided in a fair and accurate manner, although this principle does not prevent psychologists from being persuasive and forceful in the representation of their data and conclusions. Additionally, psychologists must base their assessments, recommendations, and conclusions on information and techniques that can be substantiated scientifically and that, unless it is not feasible, they should provide opinions about an individual only after examining him or her directly. Otherwise, clarification should be provided on the effect of limited information on the reliability and validity of reports and testimony. Finally, psychologists must clarify their role as consultant, expert witness, forensic examiner, or treatment provider.

Because expert testimony by mental health professionals most often involves assessment and diagnosis, additional concerns arise when formal diagnostic classification systems are used. Bloom and Rogers (1987) argued that although legal standards and principles provide the framework for developing issues that are the focus of forensic mental health evaluations, psychiatric knowledge and ethical concerns should dictate how professionals function in their role. Consequently, one important source of guidance on the use of expert testimony derives from the DSM-5 (American Psychiatric Association, 2013). Mental health professionals in the United States and much of the world rely on the DSM-5 for providing specific criteria to render diagnoses, which often form the basis for expert testimony regarding whether an individual meets the criteria for a given mental disorder. Although DSM-5 reflects the best current consensus of the psychiatric

community regarding the operational criteria for mental disorders, it has hardly been immune from scientific criticisms, such as an overly low threshold for many diagnostic categories and the inclusion of disorders of questionable validity, such as caffeine intoxication (e.g., Frances, 2013) and, in its current form, dissociative identity disorder (see Lilienfeld & Lynn, Chapter 5, this volume). Hence, it is incumbent on experts who rely on DSM-5 to acknowledge the potential limitations of DSM-5 diagnoses in their testimony.

In the remainder of the chapter, we address several topics that are often a major focus of expert testimony of clinical psychologists and other mental health professionals. In particular, we examine (1) psychological testing and assessment, (2) prediction of violent behavior, (3) eyewitness testimony, and (4) controversial psychiatric diagnoses that are occasionally introduced in the legal arena.

PSYCHOLOGICAL TESTING AND ASSESSMENT

Although legal rules of evidence provide the standards under which the admissibility of psychological assessment methods is determined, guidelines are available for selecting reliable, valid, and relevant psychological tests in forensic settings. Heilbrun (1992) noted that two types of relevance guide psychological instrument selection. The first type involves instruments that are direct measures of a specific legal question. Examples of such instruments include the Gudjonsson Suggestibility Scales as measures of interrogative suggestibility (Gudjonsson, 1997), the Rogers Criminal Responsibility Assessment Scales (Rogers, 1984), the Instruments for Assessing Understanding and Appreciation of Miranda Rights (Grisso, 1998), and the Violence Risk Appraisal Guide (Quinsey, Harris, Rice, & Cormier, 2006). The second type bears on whether an instrument measures a specific psychological construct that is merely one component of a broader legal question. Examples of this more indirect form of relevance might include measurement of parents' personality disorders in child custody evaluations (e.g., MCMI-III), intellectual and neuropsychological assessment in personal injury cases (Reynolds, 1998), and assessment of formal thought disorder in insanity evaluations (Rogers, 1984; Shapiro, 1999).

In addition, Heilburn (1992) provided seven guidelines for selecting instruments for use in forensic settings:

1. The test should be commercially available, have an adequately documented manual, and have been peer reviewed.
2. The reliability of the instrument should be established, with a coefficient (e.g., Cronbach alpha) of .80 advisable, or explicit justification for lower coefficients.

3. The test should be relevant to the legal issue, or measure a psychological construct related to the legal issue, with available validation research.
4. Test administration should be standardized.
5. The test should be applicable to the population and clinical purpose intended.
6. Structured tests and actuarial data are preferred.
7. The individual's response style (e.g., social desirability) should be considered when interpreting test results.

Another framework for assessing the admissibility of psychometric evidence is Marlowe's (1995) hybrid model, which blends scientific and legal principles. According to this model, the analysis of admissibility for psychometric evidence follows a logical course from initial questions regarding expert witness qualifications, the falsifiability and level of acceptance of data collection procedures and psychological tests, and the implementation of proper procedures for collecting and analyzing data (e.g., norms, standardized administration). In the latter stages of Marlowe's hybrid model, questions are directed at issues related to how the expert interprets test data. For instance, issues about the relevance of the test to the clinical question at hand, social implications of test use (e.g., discrimination, racial bias), and validity of the expert's conclusions must also be addressed. The expert carries the burden of being able to provide clear reasoning as to why a specific instrument was selected and how psychometric data relate to the specific psycholegal issue.

PREDICTION OF VIOLENT BEHAVIOR

A key issue in many civil and criminal cases involving involuntary commitment and future dangerousness is violence risk. For example, evidence of mental illness or mental abnormality, coupled with an imminent risk of danger to self or others, is typically required for civil commitment (*Allen v. Illinois*, 1986). Furthermore, the United States Supreme Court case of *Kansas v. Hendricks* (1997) established that it is permissible for states to civilly commit felons who have completed their sentences if evidence exists of a mental abnormality or personality disorder that makes it likely the individual will engage in predatory acts of sexual violence. In light of this controversial legal decision (Yung, 2011), courts are increasingly likely to seek out expert testimony regarding an individual's propensity for violence, which will reignite the debate over whether mental health professionals can predict violent behavior.

For many decades, some psychologists assumed that mental health professionals were no better predictors of future violence than were informed nonprofessionals. Monahan's (1981/1995) influential conclusion

that mental health professionals' predictions of violence risk are correct only one out of three times was often cited in this regard. However, much has changed in the last three decades, as (1) research has shown that violence risk predictions are often better than chance (Buchanan, 2013; Mossman, 1994), (2) the accuracy of prediction of sexual violence has improved (Rice, 1997), (3) more is known about the demographic and clinical correlates/moderators of violent behavior (Monahan & Steadman, 1994; Singh, Grann, & Fazel, 2011), and (4) researchers have developed well-validated actuarial risk indices of violence (Heilbrun, Douglas, & Yasuhara, 2009; Monahan, Steadman, Robbins, Appelbaum, Banks, et al., 2005).

Three general types of testimony are possible with respect to violence risk. One type pertains to opinions regarding a person's psychological functioning, history, and psychiatric diagnosis, without regard to potential dangerousness. Such testimony is likely to be of limited use to the courts. A second type provides an assessment of (a) the person's psychological functioning, history, and diagnosis; (b) the variables that alter his or risk for violence; and (c) a general estimate of risk over time. Third and finally, expert testimony might be proffered as a definitive statement regarding the risk of violence. For example, Edens (2006) cited an example of an expert witness who argued that because a defendant was a psychopath, he had a 100% risk of reoffending. Testimony of the latter type has a very limited scientific basis and can be viewed as pseudoscientific. In contrast, expert testimony of the second type, in which general estimates are based on research findings, is likely to be of greatest use to courts based on the helpfulness standard of both *Daubert* and FRE. Heilbrun, Philipson, Berman, and Warren (1999) found that the most commonly cited preferences among psychiatrists and psychologists for communicating violence risk are statements regarding how specific variables alter risk and the use of general estimates (e.g., low, moderate, high) of risk.

EYEWITNESS TESTIMONY

The testimony of eyewitnesses is highly probative, meaning that jurors rely heavily on observer accounts of crimes, particularly when observers are confident in their identifications (Cutler, Penrod, & Dexter, 1990; Wells, Memon, & Penrod, 2006). Although judges often instruct jurors to form their own opinions about the reliability of specific witnesses, research demonstrates that jurors often make errors regarding the veracity of eyewitness accounts. Penrod and Cutler (1995) cited research that shows jurors have difficulty in differentiating accurate from inaccurate eyewitness testimony and that witness confidence is typically a poor indicator of accuracy (see also Bothwell, Deffenbacher, & Brigham, 1987; Sauer, Brewer, Zweck, & Weber, 2010). Moreover, witness confidence is often malleable and influenced by postidentification factors such as suggestions and

leading questions. Indeed, many wrongful convictions are attributable to false eyewitness identification (Wells, 1995). For example, the work of the Innocence Project reveals that of the 310 people exonerated of crimes based on DNA evidence, approximately 75% were initially convicted largely on the basis of faulty eyewitness reports (Innocence Project, 2013).

Juror accuracy regarding eyewitness accounts can be improved by exposure to expert testimony concerning variables that affect the accuracy of eyewitness accounts. For example, eyewitness reports tend to be less accurate under poor lighting conditions, when observers identify an individual of a different race, and when a weapon is present during the crime (Wells et al., 2006). Jurors tend to make fewer errors, such as placing unwarranted credibility on witnesses, after expert testimony is presented (Penrod & Cutler, 1995).

Nevertheless, many courts are skeptical of such testimony and do not always deem it admissible. In particular, some judges have argued that expert testimony on eyewitness reports fails the "beyond the ken" test, which requires that experts provide information above and beyond most jurors' commonplace knowledge about memory (Schmechel, O'Toole, Easterly, & Loftus, 2006). In *United States v. Amador-Galvan* (1997), for example, the court upheld the exclusion of expert testimony on eyewitness identification on appeal because such testimony might confuse the jury and waste time. In addition, the court held that the testimony was suspect because it relied on abstract and incomplete data and that it ignored many known variables that impact juror decision making. In a widely publicized 2006 case involving then vice-presidential aide, Scooter Libby, who was accused of "outing" ex-CIA agent Valerie Plame, U.S. District Judge Reggie Walton refused to admit the expert testimony of University of California at Los Angeles memory researcher and psychology professor Robert Bjork. Libby's defense team had wanted Bjork to present peer-reviewed evidence documenting that the inconsistencies in Libby's story over time did not prove that he had perjured himself, because such discrepancies could just as readily be chalked up to the vicissitudes of memory. Bjork's testimony, Walton insisted, could only confirm jurors' intuitive beliefs because laypersons are already well aware of the fallibility of memory. Jurors, he wrote in his summary opinion, "do not need the guidance of a memory expert to use their 'common sense' in the understanding of how memory works" (Seidman, 2006). Yet Walton's claim runs counter to survey data demonstrating that a staggering 63% of the American public believes that human memory operates like a video camera or DVD, accurately recording all of the events we experience (Simons & Chabris, 2011; see also Simons & Chabris, 2012).

Expert testimony on eyewitness identification has also been ruled as inadmissible in other cases, with some courts citing the *Daubert* standard. For instance, in *United States v. Hall* (1999), the court ruled that such testimony would not assist the jury because it addressed issues of which the

jury was already aware and would not contribute to jurors' understanding of the case. Likewise, *United States v. Kime* (1996) questioned the scientific basis of eyewitness testimony, stating that it failed to qualify as "scientific knowledge" under *Daubert*. Other courts (e.g., *Bachman v. Leapley*, 1992) have excluded such testimony when attorneys introduced it for the sole purpose of casting doubt on the credibility of a witness.

In other cases, expert testimony on eyewitness testimony has been deemed admissible. In *People v. McDonald* (1984) such testimony was initially excluded at trial but ultimately admitted because the judge ruled that it deprived the jury of important information that would have assisted them to reach an accurate decision. Moreover, many eyewitness researchers argue that courts should be required to include expert testimony concerning eyewitness recall in all criminal trials with juries (Gross, 1999).

CONTROVERSIAL PSYCHIATRIC DIAGNOSES

Although "pop" psychology diagnoses—labels that are rampant in popular culture but that have minimal scientific backing—are endemic in the media, they have been especially prevalent in courts of law. There, they have often played a major role in legal defenses, often to the consternation of psychological scientists. For example, in late 2013, a Texas teenager convicted of killing four people and seriously injuring two others while driving under the influence was sent to rehabilitation rather than prison after a defense psychologist raised the legal defense of "affluenza," a condition ostensibly afflicting children who are so spoiled by wealth that they are unable to appreciate the implications of their actions (Abcarian, 2014). Needless to say, the scientific support for this much-parodied condition is essentially nonexistent. Yet, in the case of other pop psychology diagnoses, the evidence is more mixed. In the following section, we examine the scientific and legal status of several controversial and unvalidated psychiatric diagnoses that have been introduced as defenses in courtrooms.

Battered Woman Syndrome

Battered woman syndrome is a term that Walker (1984) used to describe the repercussions of the cyclical nature of violence in domestic relationships. By 1993, battered woman syndrome gained legal acceptance as a scientifically acceptable construct in all state courts (Blowers & Bjerregaard, 1994). In general, battered woman syndrome may be used to assist the jury in evaluating the reasonableness of a defendant's belief of being in mortal or imminent danger (e.g., *People v. Humphrey*, 1996); however, the construct is typically not permitted as a means of allowing experts to testify about a defendant's state of mind at the time of the killing (e.g., *People v. Erikson*, 1997).

This purported concept encompasses several widely accepted constructs, including learned helplessness, to explain why some battered women remain in violent relationships. Although battered woman syndrome is not a formal diagnosis, it is commonly associated with posttraumatic stress disorder (Walker, 2009). Battered woman syndrome has been the subject of expert testimony in a variety of cases, including the prosecution of accused batterers and trials of battered women who have killed their abusive partners and claimed self-defense (Blowers & Bjerregaard, 1994; Magnum, 1999). Most recently, this defense was unsuccessfully invoked in the 2013 trial of Jodi Arias, an Arizona woman who defense experts claimed killed her boyfriend in response to repeated physical abuse. This defense is typically used in trials to argue that the battered woman neither provoked the attack nor used excessive force, and that no other reasonable recourse to violence was available.

Considerable variability exists across jurisdictions regarding the admissibility of battered woman syndrome (Magnum, 1999). In most jurisdictions in which battered woman syndrome testimony has been allowed, it has been confined to a general discussion of research on the condition, including the supposed characteristics of battered women and common beliefs or misconceptions that laypersons hold about battered women (Blowers & Bjerregaard, 1994; Schuller & Vidmar, 1992). In some cases, experts offer opinions regarding whether a defendant exhibits behaviors consistent with battered woman syndrome, although testimony regarding a criminal defendant's mental state at the time of the offense is generally disallowed. Although acceptance of the scientific basis of battered woman syndrome has been debated, courts have generally been more accepting of such testimony as the body of scientific research on the topic has expanded (Russell, 2010). Still, battered woman syndrome has been cited as an example in which experts in a narrow field of study may agree, and it may therefore meet the *Frye* standard of "general acceptance" in certain courts; the evidence for the syndrome itself is not at all compelling, however (Faigman et al., 2007).

Indeed, battered woman syndrome remains controversial within the scientific and legal communities. One criticism is that the characteristics of battered woman syndrome do not appear consistently among women who have experienced long-term abuse. However, research on battered woman syndrome has generally revealed that the characteristics used to describe the syndrome are sufficiently reliable to permit its acceptance as a valid construct in legal settings. Another criticism is that the term "syndrome" potentially pathologizes the victim and underemphasizes the situational context in which domestic violence arises (Russell, 2010). Yet another criticism is that it is not altogether clear that battered woman syndrome is distinct from posttraumatic stress disorder. In light of these and other criticisms, battered woman syndrome has not been adopted as a formal diagnosis. Accordingly, many courts permit experts to testify about research

findings regarding battered woman syndrome, but often restrict them from rendering an opinion regarding whether a defendant displays this condition (Schuller & Vidmar, 1992).

Rape Trauma Syndrome

Burgess and Holmstrom (1974) introduced the concept of rape trauma syndrome to capture what they viewed as a distinctive constellation of behaviors and emotional reactions, including shame, fear, self-blame, anger, humiliation, and a chaotic lifestyle arising from rape. Like battered woman syndrome, rape trauma syndrome is often viewed as a subtype of posttraumatic stress disorder. Nevertheless, advocates of the validity of rape trauma syndrome argue that victims of rape sometimes exhibit behaviors that are seemingly inconsistent with having been traumatized. These contradictory behaviors may include returning to the scene of the rape as a gesture of defiance or refusal to surrender to fear (Stefan, 1994), continuing to have contact with the assailant following the assault (Ritchie, 1998), and delaying reporting of the assault.

Expert testimony on the responses of rape victims, and rape trauma syndrome in particular, has been admitted inconsistently across various jurisdictions. For instance, the defendant in *United States v. Smith* (1998) argued that the victim was unreliable because she failed to report the rape immediately. A prosecution expert was permitted to testify that victims of rape often do not report their assaults immediately for various reasons, including fear, guilt, or shame. This evidence was permitted only to rebut a defense claim that the victim's delay in reporting was evidence of witness unreliability. In a more liberal use of testimony on rape trauma syndrome, the court in *State v. Allewalt* (1986) permitted an expert to testify that a victim's posttraumatic stress disorder was caused by rape. Such testimony might be criticized because it went beyond the bounds of scientific knowledge and addressed questions that are reserved solely for the jury (Boeschen, Sales, & Koss, 1998). In *Henson v. State* (1989), the defense was permitted to present expert testimony on rape trauma syndrome to support its argument that a rape had not occurred. The testimony suggested that the alleged rape victim's behavior was inconsistent with rape trauma syndrome. Because the characteristics of rape trauma syndrome are not universally present in rape victims, the syndrome is not a definitive indicator of rape. Moreover, many of the supposed features of rape trauma syndrome may arise from other predisposing psychological factors (e.g., extreme anxiety, emotional instability), or stressful events other than rape (see Freckelton, 2013). Still, many therapists appear to accept the validity of rape trauma syndrome because the "underlying trauma associated with the syndrome is largely unquestioned in therapy, but is the operative issue in the courtroom" (p. 6, Fagiman & Monahan, 2009).

Sexual Addiction

Beginning largely with the widely cited work of Carnes (1983), some researchers and clinicians have compared compulsive sexual behavior with substance dependence or addiction. According to them, sexual addiction may be conceptualized as a compulsive need to engage in sexual behavior or to become sexually aroused, often accompanied by tolerance and withdrawal symptoms. Moreover, sexual addiction is often viewed as encompassing excessive sexual behavior, failed efforts to reduce such behavior, and interference with important activities (e.g., work, relationships) despite knowledge of the detrimental effects of the behavior (Kafka, 2010). Nevertheless, Gold and Hefner (1998) noted that the literature on sexual addiction consists largely of theories that are based on clinical observation, with little empirical research to support the validity of sexual addiction as a unitary construct distinct from numerous other conditions (e.g., psychopathy, impulse control disorders). Recent efforts to import the concept of sexual addiction into the DSM have met with mixed support. Specifically, the proposed diagnosis of hypersexual disorder, which captures many features of sexual addiction, was excluded from the main text of DSM-5 and has instead been placed in Section III, the portion of the manual devoted to provisional conditions meriting further research (American Psychiatric Association, 2013). The lack of consistent empirical support and scientific consensus suggests that expert testimony on sexual addiction would fail to meet legal standards for admissibility.

Despite this paucity of research support, some courts appear willing to entertain the notion, even misusing the term by deviating from its definition by Carnes and others. In *United States v. Romualdi* (1996), the defendant was charged with possession of child pornography and admitted to fantasizing about having sex with young girls, despite the fact that he had no previous criminal record and his behavior was seemingly out of character. Expert testimony, which the judge considered, suggested that the defendant might have a sexual addiction. However, the facts of the case indicate that the defendant experienced fantasies only and had not engaged in any sexual activities with children or nonconsenting partners. This pointed to a diagnosis of paraphilia rather than sexual addiction, which implies compulsive sexual behavior. Expert testimony regarding compulsive sexual behavior or disturbed patterns of sexual arousal should be based on formally recognized diagnoses in DSM-5, such as paraphilias, personality disorders, and bipolar disorder.

Homosexual Panic

In a forensic context, homosexual panic has been loosely defined as a state of rage, mixed with anxiety and tension, experienced by an individual with intense fears of being homosexual in response to a homosexual advance (Chuang & Addington, 1988). Although homosexual panic is not a formal

diagnosis in DSM-5 (American Psychiatric Association, 2013), Chuang and Addington (1988) argued that the characteristics of this state might fall under the rubric of adjustment disorder or brief reactive psychosis. Nevertheless, they noted that legal defenses of crimes of aggression based on the defense of homosexual panic should more appropriately be construed as hate crimes stemming from phobic beliefs concerning homosexuality. The authors concluded that homosexual panic does not enjoy wide acceptance in the scientific community. Still, homosexual panic has been invoked as a defense in murder cases (*People v. Milner*, 1988; *State v. Escamilla*, 1994). In *State v. Escamilla* (1994), the court did not allow expert testimony on homosexual panic, whereas in a previous case (*Parisie v. Greer*, 1983) such testimony was allowed. Recent trends suggest that the defense of homosexual panic is not likely to be deemed admissible, dovetailing with the construct's weak scientific foundation.

Black Rage

A black rage defense in criminal cases relies on the assertion that an African American defendant's conduct resulted from an uncontrollable rage induced by the experience of racism in a predominantly white society or by specific individuals who provoked the defendant (Goldklang, 1997). The defense asserts further that the rage so impaired the defendant's mental state that his or her criminal responsibility should be negated. In most cases, a defense based on the theory of black rage is used in conjunction with defenses such as insanity or diminished capacity rather than self-defense.

Although first introduced in 1846 as a legal defense in conjunction with an insanity defense (*People v. Freeman*, 1847), black rage defenses have not been particularly successful (but see Covey, 2011). Although courts may be willing to consider a defense of insanity that is based on an anger-related mental disorder, such as paranoid schizophrenia or delusional disorder (*People v. Ferguson*, 1998; *United States v. Robertson*, 1974), courts will not generally permit a defense of black rage when the defendant explicitly rejects an insanity defense. Overall, there is scant empirical support for a distinct diagnostic construct of black rage.

Road Rage

Road rage is another popular term that has received considerable attention in the media as a pattern of aggressive behavior (Sharkey, 1997). Typically, road rage involves impulsive acts of aggression committed against people whose driving incurs the wrath of the perpetrator. The victim may be a pedestrian, bicyclist, or other driver, and the offender's behavior may include rude conduct, obscene gestures, aggressive retaliatory maneuvers, physical confrontation, and property damage. Although road rage has appeared tangentially in some legal cases (e.g., *People v. Ilieveski,* 1998, in which the police ticketed an individual for driving at the speed limit

in the passing lane out of concern that the driver's behavior might trigger road rage in others), the construct appears not to have been the subject of substantial expert testimony. In one study in western Australia, researchers (Harding, Morgan, Indermaur, Ferrante, & Blagg, 1998) found that most road rage cases occurred in urban settings and involved male perpetrators and victims in which the perpetrator threatened or assaulted the victim. Moreover, the risk for involvement in road rage was greatest among young drivers who drove more frequently and often drove during afternoon rush hour. Moreover, Harding and colleagues noted that road rage and stranger violence share common features. Research also indicates a high level of co-occurrence between road rage and intermittent explosive disorder, a major impulse control disorder in the DSM (Galovski & Blanchard, 2002). Other frequently co-occurring conditions include borderline and antisocial personality disorder and substance use disorders (Sansone & Sansone, 2010). Despite attempts to ascertain the characteristics and correlates of road rage, the concept remains a behavioral descriptor rather than a formal disorder and requires further study before it can be applied in legal settings. As such, expert testimony in cases involving instances of road rage should be based on generally accepted diagnostic categories such as personality, substance use, and impulse control disorders.

Premenstrual Dysphoric Disorder

Unlike many other controversial syndromes and diagnoses, premenstrual dysphoric disorder (PMDD) has achieved recognition as a formal diagnosis in DSM-5 (American Psychiatric Association, 2013), moving from placement in an appendix for conditions for further study in DSM-IV-TR (American Psychiatric Association, 2000). PMDD is regarded as a more severe form of premenstrual syndrome and is diagnosed in DSM-5 as a distinct mood-related disturbance that requires the presence of at least 5 symptoms from among a list of 11. These 11 symptoms are extreme depressed mood; anxiety; extreme affective instability; persistent hostility; diminished interest in typical activities; concentration disturbance; fatigue; pronounced change in appetite; sense of being overwhelmed or out of control; not sleeping enough or sleeping too much; and physical symptoms, such as breast tenderness, headaches, or joint pain. Women diagnosed with PMDD experience five or more of these symptoms in most menstrual cycles during the past year, and these symptoms interfere with school, occupational, or social functioning.

Premenstrual syndrome has at times been invoked as the basis for an insanity defense in France and as a mitigating factor in English courts (Grose, 1998). Although premenstrual syndrome has been raised in a handful of cases in United States courts, including forgery, bankruptcy cases, and a felony charge of beating a 4-year-old, courts have generally not accepted testimony based on this syndrome. One notable exception is the case of Dr.

Geraldine Richter, acquitted in 1991 of a driving while intoxicated charge that also involved her attempting to kick a police officer in the groin. A gynecologist testified as an expert witness that her conduct was consistent with premenstrual syndrome, and another expert effectively challenged her blood alcohol readings (Karel, 1991). With PMDD now a formal diagnosis and a consensus reached on diagnostic criteria, triers of fact will be faced with new challenges in determining whether to accept PMDD as a defense.

Paraphilic Coercive Disorder

Preliminary drafts of DSM-III-R (American Psychiatric Association, 1987) included a proposed diagnosis of paraphilic coercive disorder, which was defined as a preoccupation with recurrent and intense sexual urges and sexually arousing fantasies involving the act of forcible sexual contact with a nonconsenting partner. Although the proposed diagnosis was withdrawn and did not appear in the DSM-IV, it now appears in Section III of DSM-5 as a condition meriting further research. This diagnosis has been highly controversial since its inception, largely because of concerns that it may medicalize and thereby excuse some cases of rape (Tosh, 2011). In contrast, advocates contend that the diagnosis helps to identify a distinct subset of violent sexual predators (Stern, 2010).

Because it is not a formal diagnosis, paraphilic coercive disorder has not been cited in legal cases. In light of recent sexually violent predator laws, however, this condition may become a focus of legal attention. Through these laws some criminal defendants who have committed an act of sexual violence have been civilly committed for treatment of a personality disorder or of some other mental disturbance that renders them a violent threat to others (e.g., *Kansas v. Hendricks*, 1997).

Codependency

Originally developed to describe a pattern of behavior in spouses of chemically dependent individuals, the term "codependency" has gained popular use in the context of expert testimony (Norwood, 1985). In general, codependency is a condition displayed by an individual, typically a spouse or an intimate partner, who relies excessively on another person (typically an individual with chemical dependence) to fulfill his or her dependency needs. Most research considers codependency to be a condition akin to a personality disorder that spurs some individuals to seek or maintain relationships with addicted individuals (Loughead, Spurlock, & Ting, 1998; Wells, Glickauf-Hughes, & Bruss, 1998). Nevertheless, critics have argued that this condition is heterogeneous (Anderson, 1994) and is difficult to distinguish from a number of allied conditions such as dependent and borderline personality disorders, in particular (Hoenigmann-Lion & Whitehead, 2007; Morgan, 1991). A search of legal databases in several large

state appellate courts revealed no cases in which expert testimony cited the concept of codependency. Accordingly, although the construct of codependency may be useful in formulating or conceptualizing cases for treatment purposes, it does not enjoy wide acceptance in legal settings.

Factitious Disorder by Proxy

The diagnosis of factitious disorder by proxy, originally called Munchausen's syndrome by proxy, was formerly included in the Appendix of DSM-IV as a condition requiring further study (American Psychiatric Association, 1994). It has now been promoted to a formal diagnosis in DSM-5 and is technically known as factitious disorder imposed on another. Factitious disorder by proxy is defined as a pattern of behaviors exhibited by a caregiver, typically the mother, who intentionally induces or causes repeated physical illness or symptoms in a child (Meadow, 1977). To distinguish this condition from malingering, the caregiver must be acting with deliberate intent to gain medical attention and not for financial or other secondary gain (see Shaw, Dayal, Hartman, & DeMaso, 2008, for a review of diagnostic issues). Thus, its application in court proceedings, particularly child abuse cases, may be encountered with greater frequency.

Despite its recognition in DSM-5, factitious disorder by proxy presents many problems from a scientific and evidentiary perspective. Mart (1999) noted that empirical grounds for factitious disorder by proxy are equivocal and limited to a relatively small number of controlled studies. Furthermore, Mart argued that most of the literature on this condition is based on clinical observations rather than empirical studies. Mart (2002) also cited research suggesting that this diagnosis is highly heterogeneous and fails to identify a distinct condition. Other critics have questioned whether factitious disorder by proxy is merely a descriptive label for a specific form of child abuse (Criddle, 2010).

Vollaro (1993) suggested that factitious disorder by proxy is underdiagnosed and may be more prevalent than originally thought, but Mart (1999) argued that the condition may in fact be overdiagnosed. One major pragmatic problem associated with making this diagnosis is that clinicians often lack direct evidence of a parent's creation of physical symptoms in the child.

Expert testimony on factitious disorder by proxy is typically sought to aid in the prosecution in cases of child abuse. In *People v. Phillips* (1981), a prosecution expert (who had not examined the defendant) addressed the question of whether the defendant's actions were consistent with those of a person with factitious disorder by proxy. Despite defense arguments at trial and later on appeal, this evidence was deemed admissible because the testimony was considered beyond the ken of jurors and would assist the trier of fact. Given that factitious disorder by proxy is now a formal diagnosis, controversies surrounding its introduction into the legal arena are likely to grow.

Neonaticide/Infanticide Syndrome

Neonaticide is defined as the killing of a newborn baby within hours of delivery, whereas infanticide is the killing of a child who is over one day old. Research has found meaningful differences between individuals, typically mothers, who kill a newborn infant and those who kill an older child (Bookwalter, 1998; Haapasalo & Petaja, 1999). In legal settings, neonaticide/infanticide syndrome is commonly linked to postpartum depression or postpartum psychosis in which the defendant attempts to prove that she was insane at the time of the killing of the child (Nonacs & Cohen, 1998). As of this writing, none of the 50 U.S. states have neonaticide or infanticide statutes (Malmquist, 2013).

Neonaticide/infanticide syndrome has been applied inconsistently in legal cases. Bookwalter (1998) noted different outcomes in three cases with similar facts, with one woman convicted of murder, another convicted of second-degree murder, and still another convicted of criminally negligent homicide. In some cases, particularly those arising in jurisdictions outside the United States, a criminal defendant may be found not guilty by reason of insanity due to a postpartum mental disorder (Haapasalo & Petaja, 1999). In *People v. Wernick* (1995), the New York State Court of Appeals upheld the exclusion of expert testimony on neonaticide syndrome in a case involving a woman who delivered her baby in a college dormitory bathroom and then asphyxiated the infant and disposed of the body. Although the defendant sought to raise an insanity defense, without evoking the term "neonaticide syndrome," the appellate court still ruled that because a *Frye* hearing was not held to determine whether this syndrome was generally accepted in the psychological community, the expert testimony was properly excluded by the trial judge.

Child Sexual Abuse Accommodation Syndrome

A final controversial syndrome that is often the subject of expert testimony is child sexual abuse accommodation syndrome (CSAAS). Like battered woman syndrome, CSAAS was formulated to explain a pattern of contradictory and misunderstood responses, in this case by child victims of sexual abuse (Summit, 1983). The characteristics of CSAAS include five general clusters of behaviors that presumably account for why some victims of sexual abuse delay reporting or retract their claims. The first two features of CSAAS include secrecy, whereby the perpetrator swears the child victim to silence and makes excuses for the failure or delay in reporting. A second characteristic is helplessness, in which the perpetrator is often someone close to or trusted by the child, thereby widening the power differential between perpetrator and child. This differential presumably leads to passive acceptance of the perpetrator's abusive behavior. The remaining three features of CSAAS involve the child victim's assumption of responsibility for the abuse, followed by conflict that may precipitate a delayed or

unconvincing disclosure. Finally, the child's resultant loss of credibility and emotional support may lead to retraction of the sexual abuse claim.

In legal settings, CSAAS has been invoked to explain seemingly inconsistent or illogical behavior on the part of child victims of sexual abuse, including delayed reporting, lack of witnesses or physical evidence, and retraction of abuse claims by the victim. Summit (1983) stated that social scientists' expert testimony may be crucial in clarifying the victim's seemingly illogical behavior and in overcoming jurors' misconceptions regarding the claims of child victims. Nevertheless, the use of CSAAS in the prosecution of child sexual abuse cases is controversial because the syndrome is often introduced not only to explain the reactions of child victims, but also to prove that sexual abuse occurred. Among the major difficulties with the validity of CSAAS is that (1) behaviors characterizing the syndrome are present in many children who have not been sexually abused (Levy, 1989), and (2) the existence of CSAAS is difficult to falsify because the defining features are contradictory. Moreover, most evidence refutes the claim that repeated denials and failures to disclose abuse are telltale signs that the abuse occurred (London, Bruck, Ceci, & Shuman, 2005). These challenges make it difficult to determine the appropriate test of admissibility of testimony based on CSAAS.

Summit (1992) responded to criticisms of CSAAS by arguing that CSAAS was never designed for the purpose of proving sexual abuse in any given case, despite the fact that it has been so applied in several cases. Levy (1989) maintained that CSAAS has not been empirically tested and is based on little more than clinical opinion. Accordingly, application of CSAAS in legal settings may be misleading and confusing to fact-finders. The case of *United States v. Bighead* (1997) illustrates one application of CSAAS testimony that was admitted as "specialized knowledge" to describe the general characteristics of victims of child sexual abuse. In this case, the victim had not reported the abuse when it allegedly occurred, and the expert offered testimony to explain why delayed reporting can be common among children who have experienced sexual abuse. The testimony offered no speculation regarding whether the victim exhibited indicators of CSAAS per se, and merely provided a context to understand the victim's behavior. Although such testimony may provide relevant information regarding the behavior of victims in child sexual abuse cases, it sidesteps the issue of whether the presence of characteristics said to be associated with CSAAS constitute proof that sexual abuse occurred.

IDENTIFYING SUITABLE AND UNSUITABLE EXPERT TESTIMONY UNDER EXTANT STANDARDS

To conclude this chapter, we offer general principles to distinguish scientifically based expert testimony from expert testimony based largely on

pseudoscience. We provide these guidelines with a major caveat: Judges remain the ultimate gatekeepers for determining what expert testimony is admissible. Nevertheless, competent forensic practice necessitates that experts appraise the scientific basis and support for their testimony.

Although considerable debate surrounds a number of controversial diagnoses (e.g., autistic spectrum disorders), DSM-5 provides experts with diagnostic criteria that have been evaluated with respect to their reliability and validity. Accordingly, we recommend that experts rely on DSM-5 and the broader scientific literature in formulating opinions they present to the court. Although most of the unsubstantiated diagnoses/syndromes we reviewed merit further research, experts have a scientific and ethical obligation to present any testimony associated with them with appropriate caveats.

Some areas of expert testimony, including certain psychological assessment methods, violence risk assessment, and eyewitness testimony, have garnered a research base sufficient to form the basis of expert opinion. Research and assessment methods that have been subjected to peer review are widely available (e.g., commercially available psychological tests with a technical manual). Methods that have been shown to be reliable and valid provide experts with a credible basis for their testimony. Although courts appear unwilling to admit expert testimony in certain cases, this unwillingness often occurs when the testimony addresses ultimate legal opinions within the province of the trier of fact. It is incumbent on expert witnesses to acknowledge the limits of their competence and the evidential bases of their opinions, defend the data on which they rely to support their conclusions, and, to the extent possible, buttress their opinions with rigorous research findings.

GLOSSARY

Codependency: Typically refers to the tendency of an individual, most often the spouse or intimate partner, to excessively rely on another person, who is often chemically dependent or abusive, for a sense of personal identity.

***Daubert* standard:** A legal standard based on the U.S. Supreme Court ruling in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), which holds that the Federal Rules of Evidence govern admissibility in federal courts. Specific criteria that a court may examine when ruling on admissibility of expert testimony include (1) whether a theory or technique can be tested, (2) peer review, (3) rate of error, and (4) general acceptance.

Expert testimony: Scientific, technical, or other specialized knowledge that will assist judges or juries to understand legal evidence or determine facts in a case. To be considered expert testimony, the subject matter must be outside the layperson's general knowledge.

Expert witness: A person whose knowledge, skill, experience, training, or education qualifies that person to provide opinions and other forms of expert testimony.

Federal Rules of Evidence (FRE): A set of legal rules that govern the admissibility of evidence in federal courts.

***Frye* test:** Also known as the "general acceptance" test; a legal standard for determining the admissibility of expert testimony that examines whether or not the scientific principle, theory, or procedures relied upon by the expert have been generally accepted in the particular field in which it belongs.

Homosexual panic: A state of rage, mixed with anxiety and tension, that is purportedly experienced by a person with latent homosexual tendencies that are aroused by homosexual advances. It has been proffered, but usually rejected by courts, as a legal defense in criminal cases.

Road rage: Impulsive acts of violence committed by a driver against another person or persons whose behavior evokes rage or anger in the driver.

Trier of fact: Either a judge or a jury. The trier of fact evaluates legal evidence, establishes facts, and renders verdicts in civil and criminal trials.

REFERENCES

Abcarian, R. (2014, January 15). Groundbreaking California measure would outlaw "affluenza" defense. *Los Angeles Times*. Available at *www.latimes.com/2014/jan15/local/la-me-ra-groundbreaking-california-law-would-outlaw-affluenza-defense-20140115*.

Allen v. Illinois, 478 U.S. 364 (1986).

Ambrogi, R. (2011). *Two more states adopt* Daubert, *bringing total to 32*. Retrieved December 12, 2013, from *www.ims-expertservices.com/blog/2011*.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. rev.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author

American Psychological Association. (2002). American Psychological Association ethical principles of psychologists and code of conduct. Retrieved June 29, 2013, from *www.apa.org/ethics/code/index.aspx*.

American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist*, *68*(1), 7–19.

Anderson, S. C. (1994). A critical analysis of the concept of codependency. *Social Work*, *39*(6), 677–685.

Bachman v. Leapley, 953 F.2d 440 (9th Cir., 1992).

Blau, T. H. (1998). *The psychologist as expert witness*. New York: Wiley.

Bloom, J. D., & Rogers, J. L. (1987). The legal basis of forensic psychiatry: Statutorily mandated psychiatric diagnoses. *American Journal of Psychiatry, 144*, 847–853.

Blowers, A. N., & Bjerregaard, B. (1994, Winter). The admissibility of expert testimony on the battered woman syndrome in homicide cases. *Journal of Psychiatry and Law*, *22*, 527–560.

Boeschen, L. E., Sales, B. D., & Koss, M. P. (1998). Rape trauma experts in the courtroom. *Psychology, Public Policy, and Law*, *4*, 414–432.

Bookwalter, B. E. (1998). Throwing the bath water out with the baby: Wrongful exclusion of expert testimony on neonaticide syndrome. *Boston University Law Review*, *78*, 1185–1210.

Borum, R. (1996). Improving the clinical practice of violence risk assessment. *American Psychologist*, *51*, 945–956.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, *72*(4), 691.

Buchanan, A. (2013). Violence risk assessment in clinical settings: Being sure about being sure. *Behavioral Sciences and the Law*, *31*, 74–80.

Burgess, A. W., & Holmstrom, L. L. (1974). Rape trauma syndrome. *American Journal of Psychiatry*, *131*(9), 981–986.

Cavanagh, E. D. (1999). Decision extends *Daubert* approach to all expert testimony. *New York State Bar Journal*, *71*(6), 9, 19–20.

Chuang, H. T., & Addington, D. (1988). Homosexual panic: A review of its concept. *Canadian Journal of Psychiatry*, *33*, 613–617.

Ciccone, J. R. (1992). Murder, insanity, and medical expert witnesses. *Archives of Neurology*, *49,* 608–611.

Committee on Ethical Guidelines for Forensic Psychologists. (2013). Specialty guidelines for forensic psychologists. *Law and Human Behavior*, *15*, 655–665.

Covey, R. (2011). Temporary insanity: The strange life and times of the perfect defense. *Boston University Law Review*, *91*, 1597–1668.

Criddle, L. (2010). Monsters in the closet: Munchausen syndrome by proxy. *Critical Care Nurse*, *30*(6), 46–55.

Cutler, B. L., & Kovera, M. B. (2011). Expert psychological testimony. *Current Directions in Psychological Science*, *20*(1), 53–57.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior*, *14*(2), 185.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S.Ct. 2786 (1993).

Edens, J. F. (2006). Unresolved controversies concerning psychopathy: Implications for clinical and forensic decision making. *Professional Psychology: Research and Practice, 37*(1), 59.

Faigman, D. L. (1995). The evidentiary status of social science under *Daubert:* Is it scientific, technical, or otherknowledge? *Psychology, Public Policy, and Law*, *1*, 960–979.

Faigman, D. L., Kaye, D. H., Saks, M. J., Sanders, J., & Cheng, E. K. (2007). *Modern scientific evidence: The law and science of expert testimony* (2006–2007 ed.). New York: Thomson-West.

Faigman, D. L., & Monahan, J. (2009). Standards of legal admissibility and their implications for psychological science. In K. S. Douglas, J. L. Skeem, & S. L. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 3–25). New York: Guilford Press.

Faigman, D. L., Saks, M. J., Sanders, J., & Cheng, E. K. (2008). Modern scientific evidence: Standards. *Statistics and Research Methods.* New York: Thomson West.

Faust, D. (2012). *Coping with psychiatric and psychological testimony* (6th ed.). New York: Oxford University Press.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241,* 31–35.

Frances, A. (2013). Is criticism of DSM-5 "Anti psychiatry?" Retrieved November 14, 2013, from *www.huffingtonpost.com/allen-frances/is-criticism-of-dsm-5-ant_b_3326980.html*.

Freckelton, I. (2013). Brain injuries and coercive care. In N. B. McSherry & I. Freckelton (Eds.), *Coercive care: Rights, law and policy* (pp. 258–280). New York: Routledge.

Fuller, K. A. (1990). Paraphilic coercive disorder. *Journal of Sex Education and Therapy, 16*, 164–171.

Galovski, T., & Blanchard, E. B. (2002). Psychological characteristics of aggressive drivers with and without intermittent explosive disorder. *Behaviour Research and Therapy*, *40*(10), 1157–1168.

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-*Daubert* world. *Law and Human Behavior*, *25*(5), 433.

General Electric Co. v. Joiner, 118 S.Ct. 512 (1997).

Gold, S. N., & Heffner, C. L. (1998). Sexual addiction: Many conceptions, minimal data. *Clinical Psychology Review*, *18*, 367–381.

Goldklang, D. L. (1997). Post-traumatic stress disorder and black rage: Clinical validity, criminal responsibility. *Virginia Journal of Social Policy and the Law*, *5*, 213–235.

Goodman-Delahunty, J. (1997). Forensic psychological expertise in the wake of *Daubert*. *Law and Human Behavior*, *21*, 121–140.

Granello, D. H., & Beamish, P. M. (1998). Reconceptualizing codependency in women: A sense of connectedness, not pathology. *Journal of Mental Health Counseling*, *20*, 344–358.

Green, E. D., & Nesson, C. R. (1992). *Federal rules of evidence: With selected legislative history and new cases and problems*. Boston: Little, Brown.

Greenberg, S. A., & Shuman, D. W. (1997). Irreconcilable conflict between therapeutic and forensic roles. *Professional Psychology: Research and Practice*, *28*, 50–57.

Grisso, T. (1998). *Instruments for assessing understanding and appreciation of Miranda rights*. Sarasota, FL: Professional Resource Press.

Grose, N. R. (1998). Premenstrual dysphoric disorder as a mitigating factor in sentencing: Following the lead of English criminal courts. *Valparaiso University Law Review*, *33*, 201–230.

Gross, W. D. (1999). The unfortunate faith: A solution to the unwarranted reliance upon eyewitness testimony. *Texas Wesleyan Law Review, 5*, 307–331.

Grudzinskas, A. J. (1999). Kuhmo Tire Col, Ltd. v. Carmichael. *Journal of the American Academy of Psychiatry and the Law*, *27*, 482–488.

Grudzinskas, A. J., & Appelbaum, K. L. (1998). General Electric Co. v. Joiner: Lighting up the post-*Daubert* landscape? *Journal of the American Academy of Psychiatry and the Law*, *26*, 497–503.

Gudjonsson, G. (1997). *The Gudjonsson suggestibility scales manual*. East Sussex, UK: Psychology Press.

Haapasalo, J., & Petaja, S. (1999). Mothers who killed or attempted to kill their child: Life circumstances, childhood abuse, and types of killing. *Violence and Victims, 14*, 219–239.

Hamilton, H. G. (1998). The movement from *Frye* to *Daubert*: Where do the states stand? *Jurimetrics*, *38*, 201–213.

Harding, R. W., Morgan, F. H., Indermaur, D., Ferrante, A. M., & Blagg, H. (1998). Road rage and the epidemiology of violence: Something old, something new. *Studies on Crime and Prevention*, *7,* 221–238.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, *16*, 257–272.

Heilbrun, K., Douglas, K. S., & Yasuhara, K. (2009). Violence risk assessment: Core controversies. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 333–357). New York: Guilford Press.

Heilbrun, K., Philipson, J., Berman, L., & Warren, J. (1999). Risk communication: Clinicians' reported approaches and perceived values. *Journal of the American Academy of Psychiatry and the Law*, *27*, 397–406.

Henson v. State, 535 N.E.2d 1189 (Ind. 1989).

Hoenigmann-Lion, N. M., & Whitehead, G. I. (2007). The relationship between codependency and borderline and dependent personality traits. *Alcoholism Treatment Quarterly*, *24*(4), 55–77.

Innocence Project. (2013). Eyewitness misidentification. Retrieved December 13, 2013, from *www.innocenceproject.org/understand/Eyewitness-Misidentification.php*.

Kafka, M. P. (2010). Hypersexual disorder: A proposed diagnosis for DSM-V. *Archives of Sexual Behavior*, *39*(2), 377–400.

Karel, R. (1991, September 6). Controversy follows DWI acquittal based on premenstrual syndrome defense. *Psychiatric News*, pp. 16–18.

Klassen, D., & O'Connor, W. (1988). A prospective study of predictors of violence in adult male mental patients. *Law and Human Behavior*, *12*, 143–158.

Klassen, D., & O'Connor, W. (1990). Assessing the risk of violence in released mental patients: A cross-validation study. *Psychological Assessment*, *1*, 75–81.

Kovera, M. B., & McAuliff, B. D. (2009). *Attorneys' evaluations of psychological science: Does evidence quality matter?* Manuscript under review.

Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology, Public Policy, and Law*, *7*(2), 267.

Kumho Tire Co., Ltd. v. Carmichael 119 S.Ct. 1167 (1999).

Landsman, S. (1995). Of witches, madmen, and products liability: An historical survey of the use of expert testimony. *Behavioral Sciences and the Law, 13*, 131–157.

Levy, R. J. (1989). Using "scientific" testimony to prove child sexual abuse. *Family Law Quarterly*, *23*, 383–409.

Littleton, R. W. (1999). Supreme Court dramatically changes the rules on experts. *New York State Bar Journal*, *71*6(8), 10–18.

London, K., Bruck, M., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse: What does the research tell us about the ways that children tell? *Psychology, Public Policy, and Law*, *11*(1), 194.

Loughead, T. A., Spurlock, V. L., & Ting, Y. (1998). Diagnostic indicators of codependence: An investigation using the MCMI-II. *Journal of Mental Health Counseling*, *20*, 64–76.

Lubit, B. W. (1998). The time has come for doing science: A call for rigorous application of *Daubert* standards for the admissibility of expert evidence in the impending silicone breast implant litigation. *New York Law School Law Review*, *42*, 147–178.

Magnum, P. F. (1999). Reconceptualizing battered woman syndrome evidence: Prosecution use of expert testimony on battering. *Boston College Third World Law Journal*, *19*, 593–624.

Malmquist, C. P. (2013). Infanticide/neonaticide: The outlier situation in the United States. *Aggression and Violent Behavior*, *18*, 399–408.

Marlowe, D. B. (1995). A hybrid decision framework for evaluating psychometric evidence. *Behavioral Sciences and the Law*, *13*, 207–228.

Mart, E. G. (1999). Problems with the diagnosis of factitious disorder by proxy in forensic settings. *American Journal of Forensic Psychology*, *17*, 69–82.

Mart, E. G. (2002). Munchausen's syndrome (factitious disorder) by proxy: A brief review of its scientific and legal status. *Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work*, *1*, 55–61.

McAuliff, B. D., Kovera, M. B., & Nunez, G. (2009). Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science? *Law and Human Behavior*, *33*(3), 247–257.

McAuliff, B. D., & Groscup, J. L. (2009). Daubert and psychological science in court: Judging validity from the bench, bar, and jury box. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 26–54). New York: Guilford Press.

McCann, J. T. (1998a). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment*, *70*, 125–144.

McCann, J. T. (1998b). *Malingering and deception in adolescents: Assessing credibility in clinical and forensic settings*. Washington, DC: American Psychological Association.

McCann, J. T., & Dyer, F. J. (1996). *Forensic assessment with the Millon inventories*. New York: Guilford Press.

McPherson, S. B. (1999). Insanity and mitigation to murder. In H. V. Hall (Ed.), *Lethal violence: A sourcebook on fatal domestic, acquaintance and stranger violence* (pp. 442–467). Boca Raton, FL: CRC Press.

Meadow, R. (1977). Munchausen syndrome by proxy: The hinterland of child abuse. *The Lancet*, *310*(8033), 343–345.

Monahan, J. (1981/1995). *The clinical prediction of violent behavior*. Northvale, NJ: Aronson.

Monahan, J., & Steadman, H. J. (Eds.). (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago Press.

Monahan, J., & Steadman, H. J. (1996). Violent storms and violent people: How meteorology can inform risk communication in mental health law. *American Psychologist*, *5*(1), 931–938.

Monahan, J., Steadman, H. J., Robbins, P. C., Appelbaum, P., Banks, S., Grisso, T., et al. (2005). An actuarial model of violence risk assessment for persons with mental disorders. *Psychiatric Services*, *56*(7), 810–815.

Morgan, J. P. (1991). What is codependency? *Journal of Clinical Psychology*, *47*(5), 720–729.

Mossman, D. (1994). Assessing prediction of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*, 783–792.

Nonacs, R., & Cohen, L. S. (1998). Postpartum mood disorders: Diagnosis and treatment guidelines. *Journal of Clinical Psychiatry*, *59*(2), 34–40.

Norwood, R. (1985). *Women who love too much: When you keep wishing and hoping he'll change*. New York: Tarcher/St. Martin's Press.
Parisie v. Greer 705 F.2d 882 (7th Cir. 1983).
Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, *1*, 817–845.
People v. Erikson, 67 Cal.Rptr. 740 (Cal. 1997).
People v. Ferguson, 248 A.D. 2d 725, 670 N.Y.S. 2d 327 (N.Y. App. Div. 1998).
People v. Freeman, 4 Denio 9 (N.Y. Sup. Ct. 1847).
People v. Humphrey, 56 Cal.Rptr. 142, 921 P.2d 1 (Cal. 1996).
People v. Ilieveski, 670 N.Y.S.2d 1004 (Monroe Co., 1998).
People v. McDonald, 37 Cal.3d 351, 690 P.2d 709 (Cal. 1984).
People v. Milner, 45 Cal.3d 227, 753 P.2d 669, (Sup. Ct. of Cal. 1988).
People v. Phillips, 122 Cal. Rptr. 703 (Cal. Ct. App. 1981).
People v. Wernick, 37 Cal. 3d 351, 690 P.2d 709 (N.Y. App. Div. 1995).
Pope, K. S., Butcher, J. N., & Seelen, J. (1993). *The MMPI, MMPI-2, and MMPI-A in court*. Washington, DC: American Psychological Association.
Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.
Reynolds, C. R. (Ed.). (1998). *Detection of malingering during head injury litigation*. New York: Plenum.
Rice, M. E. (1997). Violent offender research and implications for the criminal justice system. *American Psychologist*, *52*, 414–423.
Ritchie, E. C. (1998). Reactions to rape: A military forensic psychiatrist's perspective. *Military Medicine, 163,* 505–509.
Rogers, R. (1984). *Rogers criminal responsibility assessment scales*. Odessa, FL: Psychological Assessment Resources.
Russell, B. L. (2010). *Battered woman syndrome as a legal defense: History, effectiveness and implications*. Jefferson, NC: McFarland.
Sansone, R. A., & Sansone, L. A. (2010). Road rage: What's driving it? *Psychiatry (Edgmont)*, *7*(7), 14.
Schuller, R. A., & Vidmar, N. (1992). Battered woman syndrome evidence in the courtroom: A review of the literature. *Law and Human Behavior*, *16*, 273–291.
Seidman, J. (2006). No memory expert for Libby trial: Judge called proposed key witness a "waste of time." Politics on NBC NEWS.com. Retrieved July 5, 2013, from *www.nbcnews.com/id/15534713/ns/politics/t/no-memory-expert-libby-trial/#.UdgRA1MrOHk*.
Shapiro, D. L. (1999). *Criminal responsibility evaluations: A manual for practice*. Sarasota, FL: Professional Resource Press.
Sharkey, J. (1997, September 28). It's a mad, mad, mad, mad world: You're not bad, you're sick. It's in the book. *The New York Times*.
Shaw, R. J., Dayal, S., Hartman, J. K., & DeMaso, D. R. (2008). Factitious disorder by proxy: Pediatric condition falsification. *Harvard Review of Psychiatry*, *16*(4), 215–224.
Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the US population. *PloS One*, *6*(8), e22757.
Simons, D. J., & Chabris, C. F. (2012). Common (mis) beliefs about memory: A replication and comparison of telephone and mechanical Turk survey methods. *PloS One*, *7*(12), e51876.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499–513.

State v. Allewalt, 308 Md. 89, 517 A.2d 741 (Md. 1986).

State v. Escamilla, 245 Neb. 13, 511 N.W.2d 58, (Neb. 1994).

Steadman, H. J., McGreevy, M. A., Morrissey, J. P., Callahan, L. A., Robbins, P. C., & Cirincione, C. (1993). *Before and after Hinckley: Evaluating insanity defense reform*. New York: Guilford Press.

Steele, D. L. (1999). Expert testimony: Seeking an appropriate admissibility standard for behavioral science in child sexual abuse prosecutions. *Duke Law Journal*, *48*, 933–973.

Stefan, S. (1994). The protection racket: Rape trauma syndrome, psychiatric labeling, and law. *Northwestern University Law Review*, *88*, 1271–1345.

Steiner, M. (1997). Premenstrual syndromes. *Annual Review of Medicine, 48*, 447–455.

Stern, P. (2010). Paraphilic coercive disorder in the DSM: The right diagnosis for the right reasons. *Archives of Sexual Behavior*, *39*(6), 1443–1447.

Summit, R. C. (1983). The child sexual abuse accommodation syndrome. *Child Abuse and Neglect, 7*, 177–193.

Summit, R. C. (1992). Abuse of the child sexual abuse accommodation syndrome. *Journal of Child Sexual Abuse, 1*, 153–163.

Tosh, J. (2011). Feminist sexology and activism: Challenges to the medicalisation of sex. *British Psychological Society*, *84*, 26–29.

United States v. Amador-Galvan, 9 F.3d 1414 (9th Cir., 1997).

United States v. Bighead, 128 F.3d 1329 (9th Cir., 1997).

United States v. Frye, 293 F. 1013 (D.C. Cir., 1923).

United States v. Hall, 165 F.3d 1095 (7th Cir., 1999).

United States v. Kime, 99 F.3d 870 (8th Cir., 1996).

United States v. Robertson, 507 F.2d 1148 (D.C. Cir., 1974).

United States v. Romualdi, 101 F.3d 971 (3rd Cir., 1996).

United States v. Smith, 1998 LEXIS 5772 (6th Cir., 1998).

Vollaro, T. (1993). Muchausen syndrome by proxy and its evidentiary problems. *Hofstra Law Review, 22,* 495–520.

Walker, L. (1984). *The battered woman syndrome*. New York: Springer.

Walker, L. E. (2009). *Battered woman*. HarperCollins e-books.

Wells, G. L. (1995). Scientific study of witness memory: Implications for public and legal policy. *Psychology, Public Policy, and Law, 1,* 726–731.

Wells, M., Glickauf-Hughes, C., & Bruss, K. (1998). The relationship of co-dependency to enduring personality characteristics. *Journal of College Student Psychotherapy, 12*, 25–38.

Yung, C. (2011). Symposium: Preventative detention: Sex offender exceptionalism and reventive detention. *Journal of Criminal Law and Criminology*, *101*(3), 969–1004.

Ziskin, J., & Faust, D. (1988). *Coping with psychiatric and psychological testimony* (4th ed.). Marina del Ray, CA: Law and Psychology Press.

CHAPTER FIVE

Dissociative Identity Disorder

A Contemporary Scientific Perspective

Scott O. Lilienfeld and Steven Jay Lynn

Dissociative identity disorder (DID), known formerly as multiple personality disorder (MPD), has long been among the most controversial of all psychiatric diagnoses (see McCann, Lynn, Lilienfeld, Shindler, & Hammond, Chapter 4, this volume, for a review of other controversial psychiatric diagnoses and their legal status). The controversies surrounding DID have centered primarily on its descriptive psychopathology, diagnosis, etiology, and treatment (see also Elzinga, van Dyck, & Spinhoven, 1998; Giesbrecht, Lynn, Lilienfeld, & Merkelbach, 2008; Lynn, Lilienfeld, Merckelbach, Giesbrecht, McNally, et al., 2014). Although these controversies have a lengthy history, they have become especially divisive and even acrimonious over the past two decades. To a large extent, these debates center on fundamental questions regarding the nature, boundaries, and etiology of the diagnosis itself.

Some prominent researchers (e.g., Ross, 1997) believe that DID is one of the most commonly overlooked diagnoses in psychiatry and clinical psychology. According to these investigators, DID's prevalence and impact on psychiatric disability have been greatly underestimated (see also Dell, 2001; Dell & O'Neil, 2010). Yet surveys of clinicians indicate that many professionals are deeply skeptical of the DID diagnosis and of many prevailing theories of its etiology (Cormier & Thelen, 1998; Dell, 2001; Pope, Oliva, Hudson, Bodkin, & Gruber, 1999). These critics typically contend that DID is overdiagnosed, inadvertently created by careless mental health professionals, or both (McHugh, 2008). Some even argue that DID is a fad that enjoyed a brief stint of popularity that is now waning. For example,

Pope, Barry, Bodkin, and Hudson (2005) found that the number of publications on DID peaked in the mid-1990s and declined precipitously by 2003 (see also Paris, 2012). Because many of the points of contention surrounding DID's scientific status bear potentially important implications for the causes and treatment of other psychological conditions, they may serve as a valuable object lesson for mental health professionals.

In this chapter, we provide an overview of the major controversies regarding the scientific status of DID and, to a lesser extent, dissociative disorders in general. In addition, we attempt to outline areas of potential common ground among individuals who hold markedly differing viewpoints regarding DID and also to delineate fruitful areas for further investigation.

A BRIEF HISTORY OF DID

Early Conceptions of DID

Reports of DID in the popular and clinical literature date back at least to the 19th century. Robert Louis Stevenson's classic 1885 novel *The Strange Case of Dr. Jekyll and Mr. Hyde*, which describes the case of a scientist who ingests a mysterious potion that transforms him into an entirely different person, is among the first tales that anticipate the modern-day notion of DID.

Around the turn of the century, the French neurologist Pierre Janet (1927) introduced the concept of dissociation (or "desagregation" as he termed it), which he regarded as a means of walling off disturbing experiences from conscious awareness. For Janet, this process resulted in "double consciousness," which is similar in many ways to the modern-day concept of DID. Freud and his followers, however, were skeptical of the notion of multiple personality disorder and proposed that most or all cases of this condition were due largely to a misuse of the transference relationship, namely, the suggestive influence of therapists on patients. Freud jettisoned Janet's concept of dissociation (i.e., horizontal splitting within different parts of the unconscious) and replaced it with the concept of repression (i.e., vertical splitting between the conscious and unconscious). According to Freud, painful memories are not compartmentalized into different regions of the unconscious, as seen in the putatively distinct personalities of DID individuals, but rather they are banished into the unconscious and separated from conscious awareness.

Although the remarkable signs and symptoms of DID captured the imagination of authors and researchers throughout the 19th and 20th centuries, reports of this condition were extremely rare until the late 20th century. As of 1970, there was a total of 79 well-documented cases of DID in the world literature. Perhaps the best known early case of DID was that of "Miss Beachamp," which psychologist Morton Prince (the founder of the

Journal of Abnormal Psychology) reported around the turn of the century (Prince, 1905).

Another relatively early celebrated case of DID was that of Chris Sizemore, which formed the basis of the book (and later the Hollywood film), *The Three Faces of Eve* (Thigpen & Cleckley, 1957). Sizemore reported three personalities: Eve White, Eve Black, and a third personality named Jane. As in many cases of DID (see "Descriptive Features and Correlates of DID"), two of the personalities exhibited almost diametrically opposed personality characteristics. Eve White was reserved, traditional, and demure, whereas Eve Black was flamboyant, fun-loving, and seductive. This case attracted considerable public attention, largely because it was one of the few clear-cut cases of DID known at that time.

The DID Epidemic Begins

Beginning in the mid- to late 1970s, however, cases of DID began to be reported in substantial numbers. As of 1986, the number of reported cases had swollen to approximately 6,000. This massive increase followed closely upon the release of the best-selling book *Sybil* (Schreiber, 1973) in the mid-1970s, which told the story of a young woman (whose actual name was Shirley Ardell Mason) with 16 personalities who reported a history of severe and sadistic child abuse at the hands of her mother. The book was turned into a widely viewed and Emmy Award-winning television film in 1976 starring Sally Field.

Interestingly, however, a well-known psychiatrist who was involved closely with the Sybil case recently contended that Sybil's DID was largely or entirely the product of therapeutic suggestion. Herbert Spiegel, who served as a back-up therapist for Sybil, maintained that Sybil's primary therapist, Cornelia Wilbur, frequently encouraged her to develop and display different personalities in therapy. In addition, according to Spiegel, Wilbur referred to Sybil's personalities by different names and communicated with them individually. Spiegel further maintained that Wilbur and Flora Schreiber, who ultimately authored the best-selling book about Sybil, insisted that Sybil be described in the book as a "multiple" to make the book more appealing to the publisher (see Acocella, 1998). Indeed, in a devastating expose, journalist Debbie Nathan (2011) corroborated these claims and added fuel to the fire by contending that many of Wilbur's assertions were blatantly fabricated. As Nathan pointed out, for example, there is precious little objective evidence that Sybil's mother abused her. Nathan also documented that Wilbur repeatedly used highly suggestive techniques, including repeated prompting of alternate personality states and administration of sodium pentothal (so-called truth serum) with Sybil in an effort to bring out hidden identities and purported repressed memories of abuse (see Lynn, Krackow, Loftus, Locke, & Lilienfeld, Chapter 8, this volume, for a discussion of evidence that supposed truth serums actually boost the

risks of false memories; Piper, 1993). As we will see shortly, the role of therapeutic suggestion in Sybil's case and in other cases of DID is probably the most contentious issue in the DID literature.

The number of reported cases of DID at the turn of the 21st century is difficult to estimate, although one estimate places the number as of 1998 at approximately 40,000 (Marmer, 1998). Moreover, a number of celebrities, including comedian Roseanne Arnold and former professional football star running back Herschel Walker (see Walker, 2008), have announced that they suffer from DID, and television and film coverage of DID has skyrocketed over the past two decades (Byrne, 2001; Spanos, 1996; Showalter, 1997; Trifonova, 2010; Wilson, 2003). The reasons for the recent "epidemic" (Boor, 1982) in the number of reported DID cases of DID are still unknown. As we will see shortly, the causes of this remarkable secular increase remains a point of considerable debate among researchers and clinicians.

At least two other changes over time in the characteristics of patients with DID are worth noting. First, the number of DID personalities has increased dramatically over time. Whereas most cases of DID prior to the 1970s were characterized by only one or two personalities, recent cases are typically characterized by considerably more personalities (North, Ryall, Ricci, & Wetzel, 1993). For example, Ross, Norton, and Wozney (1989) reported that the mean number of DID personalities was 16, which was precisely the number reported by Sybil (Acocella, 1998). Second, although few individuals with DID prior to Sybil reported a history of child abuse, a substantial proportion of cases of DID that followed in the wake of Sybil reported such a history (Spanos, 1996).

DESCRIPTIVE FEATURES AND CORRELATES OF DID

Major Diagnostic Features of DID

According to DSM-5, DID is first and foremost a disorder of identity disturbance. This disturbance is manifested in profound differences across two or more independent "personality states." In turn, these personality states differ substantially from one another in their self-concept, as manifested by differences in their mood, thinking, behavior, memory, perception, and other psychological characteristics. Other dissociative disorders in DSM-5 include depersonalization/derealization disorder and dissociative amnesia. Prior to DSM-III (American Psychiatric Association, 1980), dissociative disorders were combined with somatoform disorders into a superordinate class of conditions traditionally referred to as "hysterical" disorders. Increasing research evidence, however, of differences between dissociative and somatoform conditions led to their separation in the classification system (Hyler & Spitzer, 1978).

According to DSM-5, DID is characterized by the presence of two or

more distinct "personality states." Each of these states is marked by pronounced differences in affect, cognition, perception, memory, sensory–motor functioning, and identity (American Psychiatric Association, 2013). In contrast to its predecessor, DSM-IV (American Psychiatric Association, 2000), DSM-5 now encompasses possession-related phenomena (sometimes seen in non-Western cultures as well as in Western cultures) in its criteria for DID. The evidence of multiplicity in DSM-5 may be reported by the patient, observed by others, or both. These alternate personality states or "alters" often exhibit personality features that differ markedly from those of the primary or "host" personality. In some cases, these features appear to be the exact opposite of those exhibited by the host personality. For example, if the host personality is shy and retiring, one or more of the alters may be outgoing or flamboyant. The widely publicized case of Chris Sizemore, described earlier, illustrates this phenomenon. Some therapists (e.g., Allison, 1974) have even argued that patients with DID possess an "inner self-helper," a part of the personality that is aware of everything that is occurring to alters and that can assist in their integration. Nevertheless, this hypothesis has not garnered widespread support. According to DSM-5, the presence of alters, as well as other features of DID, must not be attributable to either substance (e.g., alcohol intoxication) uses or a medical condition (e.g., temporal lobe epilepsy). DSM-5 also requires that the signs and symptoms of DID are not exclusively a transient component of cultural or religious practices, such as glossolalia (speaking in tongues) which occurs during religious ceremonies.

In addition, according to DSM-5, individuals with DID must report substantial gaps in memory for ordinary events, important information about themselves, traumatic events, or all three. For example, they may report frequent periods of "lost time" lasting hours or days in which they cannot recall where they were or what they were doing. This amnesia is often reported to be asymmetrical, whereby the host personality knows little about the behaviors of the alters, but not vice versa (American Psychiatric Association, 2000).

Nevertheless, the scientific standing of amnesia as a feature of DID is controversial. Allen and Iacono (2001) concluded that controlled laboratory studies examining the transfer of explicit and implicit memories offer minimal support for the claim that patients with DID actually experience amnesia across alters (but see Dorahy, 2001, for a somewhat different conclusion). For example, researchers have found little or no evidence for interidentity amnesia when using objective measures (e.g., behavioral tasks or event- related potentials) of memory (Giesbrecht et al., 2010; Merckelbach, Devilly, & Rassin, 2002). In one recent study, Huntjens, Verschuere, and McNally (2012) used a concealed information task and found clear evidence of transfer of autobiographical memory (i.e., memories of childhood sexual abuse) across alters. These findings call into question the assumption that DID alters harbor memories that are insulated from each other by amnestic barriers.

In addition, research by Read and his colleagues (e.g., Belli, Winkielman, Read, Schwarz, & Lynn, 1998; Read & Lindsay, 2000) demonstrates that one can readily induce reports of autobiographical memory gaps in normal subjects simply by asking them to recall multiple events from early childhood. Specifically, individuals who are asked to recall multiple events from early childhood (as often occurs in depth-oriented psychotherapy) will typically do so obligingly. As a consequence, when they are asked such questions as "Was there ever a period of time when you remembered less of your childhood than you do now?," they will typically respond "Yes" because they are accurately reporting that they now recall (or at least believe that they recall) more of their childhood history than they once did. In fact, these and similar questions are commonly used in investigations of DID to verify the presence of amnesia (see Ross, 1997). Self-reports of autobiographical memory gaps in patients with DID must therefore be interpreted with caution, particularly when patients have been asked repeatedly to recall childhood memories.

Demographic and Familial Correlates of DID

Relatively little is known about the demographic or familial correlates of DID. For example, the prevalence of DID in the general population is controversial. Until fairly recently, it was widely assumed that DID is exceedingly uncommon. DSM-III (American Psychiatric Association, 1980), for example, stated that MPD, as it was then called, "is apparently extremely rare" (p. 258). Nevertheless, DSM-IV (American Psychiatric Association, 2000) was conspicuously silent regarding the prevalence of DID and noted only that reports of its prevalence have been highly variable across studies. DSM-5 cites data indicating that the prevalence of DID in the general population may be as high as 1.5% (American Psychiatric Association, 2013, p. 294). Indeed, although some authors (e.g., Paris, 2012; Piper, 1997) claim that genuine DID is either nonexistent or very rare (see also Rifkin, Ghisalbert, Dimatou, Jin, & Sethi, 1998), other authors (e.g., Ross, 1997) maintain that DID is at least as common as schizophrenia. For example, presaging the text of DSM-5, Ross (1997) estimated that between 1 and 2% of the North American population meets the criteria for DID. These discrepancies among authors are difficult to resolve given the absence of clear-cut external validating variables (see Robins & Guze, 1970) for DID.[1]

[1] One important controversy regarding DID that we touch on only briefly in this chapter is the question of whether this condition is overdiagnosed using structured interviews (see Elzinga et al., 1998). This issue is extremely difficult to settle at the present time owing to the absence of dependable external validating "criteria" for the presence or absence of DID. See Elzinga et al. (1988), Gleaves (1996), Lilienfeld et al. (1999), and Ross (1991) for discussions of this controversy.

Most early prevalence studies showed a marked female predominance, with most sex ratios ranging from 3 to 1 to 9 to 1 across clinical samples (American Psychiatric Association, 2000), although DSM-5 asserts that the sex difference in community samples is minimal or absent. Some authors argue that the imbalanced sex ratio found in many early studies may have been an artifact of selection and referral biases, and that a large proportion of males with DID end up in prisons (or other forensic settings) rather than in clinical settings (Putnam & Loewenstein, 2000). In general, women with DID tend to report more alters than do men (American Psychiatric Association, 2000).

The results of several controlled studies indicate that DID co-aggregates within biological families (American Psychiatric Association, 2013). Nevertheless, the absence of any twin or adoption studies of DID or other dissociative disorders precludes us from ascertaining the extent to which such familial clustering is due to genes, shared environment, or both (but see Jang, Paris, Zweig-Frank, & Livesley, 1998; Waller & Ross, 1997, for data on genetic and environmental influences on trait dissociation).

DID Alters

The nature and features of DID alters are highly variable both across and within individuals. The number of alters has been reported to range from one (the so-called split personality) to hundreds or even thousands. One clinician reported a case of a DID patient with 4,500 alters (Acocella, 1998). These alters are not uncommonly of different sexes, ages, and even races. There have even been reported alters of Mr. Spock, Teenage Mutant Ninja Turtles, lobsters, chickens, gorillas, tigers, unicorns, panthers, God, the bride of Satan, and the rock star, Madonna (Acocella, 1998; Ganaway, 1989; Piper & Merskey, 2004).

Some of the reported differences among alters have been striking. For example, alters have been reported to differ in their allergies, handwriting, voice patterns, eyeglass prescriptions, handedness, and other psychological and physical characteristics. Frank Putnam, a major researcher on DID, even reported a case of DID in which one alter, but not other alters, exhibited cardiac arrhythmia (*The Infinite Mind*, 1998).

Nevertheless, virtually all of these reported differences derive from anecdotal and uncontrolled reports. Moreover, most of these reports have not controlled adequately for naturally occurring variability in these characteristics over time. Both handwriting and voice, for example, often show at least some variability over time within individuals, especially in response to situational variables (e.g., fatigue, stress), and some allergies have been demonstrated to be susceptible to classical conditioning. As a consequence, these and other reported differences across alters are difficult to interpret with confidence (see also Spanos, 1996, for a critique).

Several researchers have also reported psychophysiological differences

across alters. For example, investigators have reported differences among alters in respiration rate (e.g., Bahnson & Smith, 1975), electroencephalographic (brain wave) activity (e.g., Ludwig, Brandsma, Wilbur, Bendfeldt, & Jameson, 1972), and skin conductance responses (e.g., Brende, 1984). Nevertheless, these and other psychophysiological differences (see also Putnam, Zahn, & Post, 1990) do not provide especially compelling evidence for the existence of qualitatively distinct differences among alters. As Allen and Movius (2000) noted, such differences could be attributable to changes in mood or cognition over time or to temporal changes in variables (e.g., levels of muscle tension) that are largely under volitional control. Moreover, at least some of these differences may be attributable to Type I error, given the large number of psychophysiological variables examined in many of these investigations (Allen & Movius, 2000).

One approach to addressing these criticisms is the use of simulator designs in which investigators ask non-DID individuals to simulate (mimic) or role-play DID alters. In most studies, comparisons between patients with DID and DID simulators have not revealed significant differences on measures of memory, event-related potentials, or self-reported dissociative experiences (Boysen & VanBergen, 2013). In contrast, in a recent interesting study, Reinders, Willemsen, Vos, den Boer, and Nijenhuis (2012) asked non-DID individuals with high levels of fantasy-proneness to simulate DID personality states while responding to scripts of either past traumatic or neutral experiences. Using measures of cerebral regional blood flow, they reported differences between simulating participants and individuals with DID in response to the aversive memories. Specifically, Reinders et al. found that, in contrast to simulated personality states, DID personality states displayed different patterns of brain activation across traumatic and neutral conditions.

Nevertheless, these intriguing findings are somewhat different to interpret. Although individuals with DID may display marked psychophysiological responses to trauma-related stimuli compared with non-DID simulators, such differences may merely reflect the fact that traumatic memories are far more emotionally impactful for individuals with DID than for simulators who are simply enacting identities. Hence, these findings do not provide definitive evidence for or against the contention that DID alters represent distinct identity states.

The "Multiple Personalities" Controversy

One longstanding controversy concerns the question of whether individuals with DID harbor qualitatively distinct "personalities," each with its own unique pattern of life experiences, personality traits, interests, and attitudes. Some authors, such as Braun (1986), maintain that patients with DID do indeed possess separate personalities in addition to "fragments," that is, aspects of personalities. Indeed, the older term "multiple personality

disorder" in DSM-III and DSM-III-R clearly implies the existence of unique and largely independent cohabiting personalities.

Many advocates of the DID diagnosis now argue that DID is not characterized by the presence of independent and fully developed personalities (Ross, 1990, 1997). Coons (1984), for example, contended that "it is a mistake to consider each personality totally separate, whole, or autonomous. The other personalities might best be described as personality states, other selves, or personality fragments" (p. 53). Ross (1994) similarly asserted that "much of the skepticism about MPD is based on the erroneous assumption that such patients have more than one personality, which is, in fact, impossible" (p. 81). David Spiegel (1993), who was chair of the DSM-IV task force on dissociative disorders, wrote that "there is a widespread misunderstanding of the essential psychopathology in this dissociative disorder, which is a failure of integration of various aspects of identity, memory, and consciousness. The problem is not having more than one personality; it is having less than one personality" (p. 15). In recognition of this point, DSM-5 jettisoned wording from DSM-IV implying that many individuals with DID harbor distinct personalities, instead substituting the phrase "personality states" (American Psychiatric Association, 2013, p. 292). Moreover, DSM-5 removed the requirement in DSM-IV that alter identities take recurrent control over the person's behavior.

Still, the ongoing question of whether patients with DID possess distinct coexisting personalities is of more than semantic significance. For example, in legal cases questions have arisen concerning whether individuals with DID should be held criminally responsible if one of their alter personalities committed a crime or whether each alter personality is entitled to separate legal representation. Some attorneys have invoked DID as an insanity defense, asserting that one or more of the alters, rather than the host personality, committed the crime in question (Farrell, 2011). Some trial judges have even required that all DID personalities be sworn in before providing testimony (Slovenko, 1999). In addition, if patients with DID truly possess independent and fully developed personalities, this would pose significant challenges to models of the DID's etiology. For example, how do these ostensibly complete personalities, each presumably with its own set of personality traits and attitudes, form? For patients who possess hundreds of alters, is each personality genuinely independent of the others, or are certain personalities merely variants or slightly different manifestations of the others?

THE ETIOLOGY OF DID: TWO COMPETING MODELS

DID's "Existence": A Pseudocontroversy

The principal controversy regarding the scientific status of DID has often been framed in terms of whether this condition "exists" (e.g., Arrigo &

Pezdek, 1998; Dunn, Paolo, Ryan, & van Fleet, 1994; Mai, 1995; see also Hacking, 1995). Nevertheless, as we and our colleagues have argued elsewhere (Lilienfeld et al., 1999), the question of DID's "existence" is a pseudocontroversy. There is little dispute that DID "exists," in that a number of individuals exhibit multiple identity enactments (i.e., apparent alters) in conjunction with reported autobiographical memory gaps in childhood or adolescence. This point was aptly put by McHugh (1993): "Students often ask me whether multiple personality disorder (MPD) really exists. I usually reply that the symptoms attributed to it are as genuine as hysterical paralysis and seizures" (p. 4). Somatoform conditions, like DID, are unquestionably genuine, although their origins remain largely obscure.

The central question at stake therefore is not DID's existence but rather its etiology. As we will learn shortly, some researchers contend that DID is a spontaneously occurring consequence of childhood trauma, whereas others contend that it emerges primarily in response to suggestive therapist cueing, media influences, and broader sociocultural expectations. But even these skeptical researchers believe that DID is "genuine" in the sense that its signs and symptoms are typically not faked or intentionally produced.

There is general agreement, however, that at least some individuals have successfully pretended to have DID (Farrell, 2011; Merten & Merckelbach, 2013). For example, Kenneth Bianchi, one of the Hillside Strangler murderers, is widely believed to have faked DID to escape criminal responsibility (Orne, Dinges, & Orne, 1984). Nevertheless, outside of criminal settings, cases of malingered DID are believed to be quite rare, and both proponents and skeptics of the DID diagnosis agree that the substantial majority of individuals with this condition are not intentionally producing their symptoms (see Boon & Drajer, 1993, for a discussion of the problem of intentionally produced DID).

The Central Controversy: Two Competing Etiological Models

In general, two major competing views regarding the etiology of DID have emerged (see Gleaves, 1996; Lynn, Lilienfeld, Merckelbach, Giesbrecht, McNally, et al., 2014): the posttraumatic model (PTM) and the sociocognitive model (SCM). Although these two models are not mutually exclusive, they differ substantially in emphasis concerning the causes of DID. To oversimplify these views slightly, the PTD model posits that core DID features, particularly alters, are *discovered* by therapists, whereas the SCM model posits that these features are *created* by therapists. Because we believe that the bulk of the research evidence supports the SCM, we devote much of the remainder of the chapter to a discussion of this model. At the same time, we believe that certain aspects of the PTM have yet to be convincingly falsified, and therefore this model requires additional investigation. Moreover, we believe that a meaningful rapprochement between at least certain aspects of these two models may ultimately prove possible.

The Posttraumatic Model

Proponents of the PTM (e.g., Dell, 2006; Gleaves, 1996; Gleaves, May, & Cardena, 2001; Ross, 1997) posit that DID is a posttraumatic condition that arises primarily from a history of severe physical and/or sexual abuse in childhood. They typically argue that individuals who undergo horrific trauma in early life often dissociate or compartmentalize their personalities into discrete alters as a means of coping with the intense emotional pain of this trauma (Dalenberg et al., 2012). According to Ross (1997), "MPD is a little girl imagining that the abuse is happening to someone else" (p. 59). In support of this assertion, proponents of the PTM cite data suggesting that a large proportion—perhap. 90% or more—of individuals with DID report a history of severe child abuse (Gleaves, 1996). Another, more indirect, source of evidence for the PTM derives from structural brain imaging data demonstrating that the hippocampi of DID appear to be smaller than those of healthy comparison participants (Vermetten, Schmahl, Lindner, Loewenstein, & Bremner, 2006). This finding is broadly consistent with evidence from animal studies that severe stress, including that induced by abuse, may produce hippocampal damage (Bremner, 1999). Nevertheless, it is unknown whether this smaller hippocampal size preceded or followed the onset of DID in participants.

The essence of the PTM has been well articulated by philosopher Daniel Dennett (1991):

> The evidence is now voluminous that there are not a handful or a hundred but thousands of cases of MPD diagnosed today, and it almost invariably owes its existence to prolonged early childhood abuse, usually sexual, and of sickening severity. . . . These children have often been kept in such extraordinarily terrifying and confusing circumstances that I am more amazed that they survive psychologically at all than I am that they manage to preserve themselves by a desperate redrawing of their boundaries. What they do, when confronted with overwhelming conflict and pain, is this: They "leave." They create a boundary that the horror doesn't happen to them; it either happens to no one, or to some other self. (p. 150)

Proponents of the PTM attribute the dramatic increase in the reported prevalence of DID over the past few decades to the heightened awareness and recognition of this condition by psychotherapists. Specifically, they maintain that clinicians have increasingly become attuned to the presence of possible DID in their clients and as a consequence inquire more actively about potential symptoms of this condition (Gleaves, 1996). They also point out that a number of conditions, such as posttraumatic stress disorder (PTSD) and obsessive–compulsive disorder, were apparently underdiagnosed in previous decades (e.g., Zohar, 1998) and that a relatively abrupt massive increase (as occurred with DID at least through the mid-1990s) in

the reported prevalence of a condition does not necessarily call into question its validity. In many cases, the proponents of the PTM advocate the use of hypnosis, sodium amytal, or sodium pentothal (see Lynn, Krackow, Loftus, Locke, & Lilienfeld, Chapter 8, this volume), guided imagery, and other suggestive therapeutic techniques to call forth alters that have otherwise been inaccessible, as well as to recover apparently repressed memories of child abuse.

The Sociocognitive Model

In contrast to advocates of the PTM, proponents of the SCM (Spanos, 1994, 1996; see also Aldridge-Morris, 1989; Lilienfeld et al., 1999; Lynn et al., 2012a; Lynn et al., 2012b; Lynn & Pintar, 1997; McHugh, 1993, 2008; Merskey, 1992; Sarbin, 1995) contend that DID is largely a socially constructed condition that results from inadvertent therapist cueing (e.g., suggestive questioning regarding the existence of possible alters), media influences (e.g., film and television portrayals of DID), and broader sociocultural expectations regarding the presumed clinical features of DID. For example, proponents of the SCM believe that the release of the book and film *Sybil* in the 1970s played a substantial role in shaping conceptions of DID in the minds of both of the general public and psychotherapists, and in inadvertently encouraging individuals to adopt the core features of this condition (Paris, 2012). According to Spanos (1996), Sybil "became a model of the MPD survivor that greatly influenced the expectations of therapists and patients alike" (p. 267). Interestingly, as noted earlier, reported cases of child abuse in patients with DID became widespread only following the release of *Sybil*.

Spanos (1994) and other proponents of the SCM (McHugh, 2008) contend that individuals with DID are engaged in a form of unconscious "role playing" that is similar in some ways to the intense sense of imaginative involvement that some actors report when playing a part. Because individuals who engage in role playing essentially "lose themselves" in the enacted part, this phenomenon should not be confused with simulation or conscious deception. Some authors have erroneously assumed that the SCM posits that individuals with DID are intentionally producing these features. But the SCM is careful to distinguish role playing from simulation (Lilienfeld et al. 1999; in contrast, see Gleaves, 1996).

According to the SCM, the dramatic "epidemic" in cases if DID observed in recent decades stems largely from iatrogenic (therapist-induced) influences and the increased media attention accorded to DID. Specifically, according to the SCM, as DID has become more familiar to both psychotherapists and the general public, an autocatalytic feedback loop (Hacking, 1995; see Shermer, 1997, for examples) has been set in motion. In this feedback loop, therapeutic and societal expectations regarding the features of DID have given rise to greater numbers of cases of DID, in turn influencing

therapeutic and societal expectations regarding the features of DID, in turn giving rise to a greater number of cases of DID, and so on. It is critical to emphasize that the SCM does not contend that DID is *entirely* iatrogenic because media influences and broader sociocultural expectations often play an important role in the genesis of DID. The notion that the SCM posits that DID is entirely iatrogenic represents another frequent misconception concerning this model. For example, Gleaves et al. (2001) referred to the SCM as the "iatrogenic" theory of DID (see Brown, Frischholz, & Scheflin, 1999; and Gleaves, 1996, for other examples).

Another important brick in the edifice of the SCM is the assumption that DID is merely one variant of a much broader constellation of conditions characterized by multiple identity enactments, including cases of purported demonic possession, channeling, mass hysteria, transvestism, and glossolalia that traverse cultural and historical boundaries (Spanos, 1996). From this perspective, DID is not a unique condition but is instead a superficially different manifestation of the same diathesis that gives rise to many other conditions marked by dramatically different behaviors over time, cultures, and situations. Although the protean manifestations of these role enactments are shaped by cultural and historical expectations, their underlying commonalities are suggestive of a shared etiology (Lilienfeld et al., 1999; see also Hacking, 1995).

Some proponents of the SCM (e.g., Spanos, 1994, 1996) have placed more emphasis on social-role expectations and iatrogenic influences than on individual difference variables. Nevertheless, the SCM is entirely compatible with the possibility that individual differences in certain personality traits, such as proneness to fantasy (Giesbrecht et al., 2008; Lynn, Rhue, & Green, 1988) or absorption (Tellegen & Atkinson, 1974), render certain individuals especially susceptible to suggestive therapeutic, media, and cultural influences (Lynn et al., 2012). In addition, this model is consistent with findings indicating that a substantial proportion of patients with DID meet criteria for borderline personality disorder (BPD) and other psychiatric conditions marked by unstable and unpredictable behavior, such as bipolar disorder (Ganaway, 1995; Lilienfeld et al., 1999; Lynn et al., 2011). For example, clients with BPD—who typically exhibit severe disturbances of identity, dramatic mood swings, sudden changes in feelings toward other people, and impulsive and seemingly inexplicable behaviors (e.g., self-mutilation)—may often be seeking an explanation for these puzzling symptoms, as may their therapists. Therapists who repeatedly ask such questions as "Is it possible that there is another part of you with whom I haven't yet spoken?" may gradually begin to elicit previously "latent alters" that ostensibly account for their clients' otherwise enigmatic behaviors.

Many of the key features of the SCM were nicely summed up by Frances and First (1998), who ironically were two of the principal architects (chairperson and editor, respectively) of DSM-IV, which had endorsed the traditional view of DID as a condition marked by multiple indwelling identities:

> Dissociative Identity Disorder . . . is a fascinating condition. Perhaps too much so. The idea that people can have distinct, autonomous, and rapidly alternating personalities has captured the attention of the general public, of some therapists, and of hordes of patients. As a result, especially in the United States, there has been a marked increase in the diagnosis of Dissociative Identity Disorder. Much of the excitement followed the appearance of books and movies (like *Sybil* and *The Three Faces of Eve*) and the exploitation of the diagnosis by enthusiastic TV talk show guests. . . . Many therapists feel that the popularity of Dissociative Identity Disorder represents a kind of social contagion. It is not so much that there are lots of personalities as that there are lots of people and lots of therapists who are very suggestible and willing to climb onto the bandwagon of this new fad diagnosis. As the idea of multiple personality pervades our popular culture, suggestible people coping with a chaotic current life and a severely traumatic past express discomfort and avoid responsibility by uncovering "hidden personalities" and giving each of them a voice. This is especially likely when there is a zealous therapist who finds multiple personality a fascinating topic of discussion and exploration. (pp. 286–287)

Advocates of the SCM have invoked a wide variety of pieces of research evidence in support of this theoretical position (see Lilienfeld et al., 1999; McHugh, 2008; Piper & Merskey, 2004; Spanos, 1994, 1996). In the following section, we present the major sources of evidence consistent with the SCM and examine common criticisms of the SCM by proponents of the PTM.

EVIDENCE FOR THE SOCIOCOGNITIVE MODEL OF DID

Recommended Treatment Practices for DID

One important source of evidence in favor of the SCM is the mode of treatment practices employed by some advocates of the PTM. Claims by a number of proponents of the PTM to the contrary (Brand et al., 2012; Brown et al., 1999; Gleaves, 1996), many standard therapeutic practices for DID—especially those performed by certain PTM advocates—are geared toward encouraging the appearance of alters and treating them as though they were distinct identities.

Indeed, inspection of the mainstream DID treatment literature reveals that therapists are often encouraged to reify the existence of multiple identities by mapping the system of alters and to establish direct contact with alters if they are not otherwise forthcoming (Piper, 1997). These reifying techniques are especially common in the early stages of psychotherapy, although the later stages often focus on achieving integration among alters (Ross, 1997).

For example, Kluft (1993) argued that "when information suggestive

of MPD is available, but an alter has not emerged spontaneously, asking to meet an alter directly is an increasingly accepted intervention" (p. 29). Kluft further acknowledged that his most frequent hypnotic instruction to patients with DID is "Everybody listen" (see Ganaway, 1995). Braun (1980) wrote that "after inducing hypnosis, the therapist asks the patient 'if there is another thought process, part of the mind, part, person or force that exists in the body' " (p. 213). Bliss (1980) noted that in the treatment of DID "alter egos are summoned, and usually asked to speak freely. . . . When they appear, the subject is asked to listen. [The subject] is then introduced to some of the personalities" (p. 1393). Putnam (1989) suggested using a technique known as the "bulletin board," which allows patients with DID to have a "place where personalities can 'post' messages to each other. . . . I suggest that the patient buy a small notebook in which personalities may write messages to each other" (p. 154). Ross (1997) and other therapists (e.g., Putnam, 1989) have recommended giving names to each alter in order to "'crystallize' it and make it more distinct" (p. 311). Ross (1997) also advocated the use of "inner board meetings" as "a good way to map the system, resolve issues, and recover memories" (p. 350). He described this technique as follows:

> The patient relaxes with a brief hypnotic induction, and the host personality walks into the boardroom. The patient is instructed that there will be one chair for every personality in the system. . . . Often there are empty chairs because some alters are not ready to enter therapy. The empty chairs provide useful information, and those present can be asked what they know about the missing people. (p. 351)

In addition, one increasingly popular therapeutic method, internal family systems therapy, is premised on the notion that the mind houses separate subpersonalities (e.g., protectors, firefighters, exiles) that must be accessed and integrated for healing to occur (Goulding & Schwartz, 2002; see also Pignotti & Thyer, Chapter 7, this volume).

These and other treatment recommendations derived from the mainstream DID literature (see Piper, 1997, pp. 61–68, for additional examples) strongly suggest that many therapists are explicitly encouraged to reify the existence of alters by acknowledging and validating their independent existence. Even the slightly more cautious guidelines issued recently by the International Society for the Study of Trauma and Dissociation (ISSTD) inform therapists that "in times of repeated acting out by the patient, and/or at times of therapeutic impasse, it can be essential to directly elicit or make contact with alternate identities previously known or not, that are related to these difficulties" (ISSTD, 2011, p. 140). From a behavioral or social learning perspective, the process of attending to and reifying alters may adventitiously reinforce patients' displays of multiplicity.

Another treatment practice that may inadvertently facilitate the

emergence of alters is hypnosis. Clinicians who treat patients with DID frequently use hypnosis in an effort to discover or call forth presumed latent alters (Spanos, 1994, 1996). The evidence regarding the use of hypnosis in such patients provides mixed support for the SCM. On the one hand, the results of several studies reveal few or no differences in the diagnostic features (e.g., alters, number of DID criteria) of patients with DID who have and have not been hypnotized (e.g., Putnam, Guroff, Silberman, Barban, & Post, 1986; Ross & Norton, 1989; see Gleaves, 1996, for a review). In addition, several studies indicate that many or most patients with DID have never been hypnotized (Gleaves, 1996), a finding that strongly suggests that hypnosis is not necessary for the emergence of DID.

In contrast, the finding that hypnotized and nonhypnotized patients with DID do not differ significantly in many characteristics (e.g., number of DID criteria) is difficult to interpret in light of ceiling effects (Lilienfeld et al., 1999; Powell & Gee, 1999). Specifically, given that almost all of the patients in these studies met the criteria for DID according to various diagnostic criterion sets (e.g., DSM-III), the differences in the number of DID criteria between hypnotized and nonhypnotized patients are not surprising.

In addition, in a reanalysis of the dataset of Ross and Norton (1989), Powell and Gee (1999) found that hypnotized patients exhibited greater variance in the number of alters at the time of diagnosis and in later treatment. Although the meaning of this finding is not entirely clear, it may reflect bimodal attitudes toward iatrogenesis among practitioners who use hypnosis, with some (who believe that hypnosis is potentially iatrogenic) using hypnosis never or rarely and others (who believe that hypnosis is not iatrogenic) using hypnosis frequently. Powell and Gee (1999) also found that clinicians who used hypnosis reported a significantly higher number of patients with DID in their caseloads than did practitioners who did not use hypnosis. Although this finding is open to several interpretations (e.g., DID specialists may be more likely to use hypnosis than are other clinicians), it is consistent with iatrogenesis.

Moreover, the SCM does not posit that hypnosis is necessary for the creation of DID alters. Hypnotic procedures do not possess any inherent or unique features that are necessary to facilitate responsivity to suggestion (Spanos & Chaves, 1989). Other methods, such as suggestive and leading questions, may be equally likely to induce clients' adoption of multiple identities (Barber, 1979; Spanos, 1996).

None of this implies, of course, that all or even most treatment for DID is ineffective or harmful. Naturalistic data indicate that DID often remits following treatment (Brand, Classen, McNary, & Zaveri, 2009), raising the possibility that certain DID interventions are effective. Nevertheless, studies do not permit an evaluation of the extent to which symptom reduction in dissociative patients in naturalistic studies is due to regression to the mean, the passage of time, placebo effects, or other artifacts. Other methodological limitations in treatment studies of DID include variability in treatments offered to patients (e.g., Choe & Kluft, 1995), lack of controls

for nonspecific effects (e.g., Ellason & Ross, 1997), dropout rates as high as 68% (Gantt & Tinnin, 2007), and the failure to document clinically meaningful changes following treatment. Because there are no randomized controlled trials of DID treatment, it is unknown which, if any, extant treatments are effective for DID.

The Clinical Features of Patients with DID Before and After Psychotherapy

There is compelling evidence that a large proportion—perhaps even a substantial majority—of patients with DID exhibit very few or no unambiguous signs of this condition (e.g., alters) prior to psychotherapy. For example, Kluft (1991) estimated that only 20% of patients with DID exhibit unambiguous signs of this condition and that the remaining 80% exhibit only transient "windows of diagnosability," that is, short-lived periods during which the core features of DID are observable. Virtually all authors in this literature agree that a large proportion, and perhaps a majority, of patients with DID exhibit few or no clear-cut signs of this condition prior to psychotherapy (Kluft, 1984; Putnam, Guroff, Silberman, Barban, & Post, 1986; Ross, 1997). Moreover, individuals with DID typically are in treatment for an average of 6 to 7 years before being diagnosed with this condition (Gleaves, 1996). Such evidence raises the possibility that these patients often develop unambiguous features of DID only after receiving psychotherapy.

Moreover, although systematic data are lacking, the DID literature shows general agreement that many or most patients with DID are unaware of the existence of their alters prior to psychotherapy. For example, Putnam (1989) estimated that 80% of patients with DID possess no knowledge of their alters before entering treatment, and Dell and Eisenhower (1990) reported that all 11 of their adolescent patients with DID had no awareness of their alters at the time of diagnosis. Similarly, Lewis, Yeager, Swica, Pincus, and Lewis (1997) reported that none of the 12 murderers with DID in their sample reported any awareness of their alters.

Some authors have also reported that the number of DID alters tends to increase over the course of treatment (Kluft, 1984; Ross et al., 1989). In addition, although the number of alters per DID case at the time of initial diagnosis has remained roughly constant over time (Ross et al., 1989), the number of alters per DID case in treatment has increased over time (North et al., 1993).

These findings are consistent with the SCM, as they suggest that many psychotherapeutic practices for DID may inadvertently encourage the emergence of new alters. Moreover, as we noted elsewhere (Lilienfeld et al., 1999, p. 512), one would be hard-pressed to find another DSM-5 disorder whose principal psychopathological feature (i.e., alters) is typically unobservable prior to standard treatment and becomes substantially more florid following this treatment.

At the same time, some proponents of the PTM argue that these findings are potentially consistent with this model. Specifically, they maintain that alters were merely "latent" at the time of initial diagnosis and became observable only after prompting and elicitation by therapists (e.g., Gleaves, 1996). Without independent evidence of these alters, however, this position raises serious concerns regarding the falsifiability of the PTM. That is, if the number of alters either decreased or remained constant over the course of therapy, proponents of the PTM could maintain that psychotherapy for DID either ameliorated the symptoms of the condition or successfully held potential deterioration at bay. In contrast, the finding that the number of alters tends to increase over the course of therapy has been interpreted by proponents of the PTM as indicating that psychotherapy successfully uncovered alters that were merely latent (Gleaves, 1996). Because a theoretical model that is consistent with any potential set of observations is difficult or impossible to falsify and is therefore of questionable scientifically utility (Popper, 1959), proponents of the PTM will need to make explicit what types of evidence could falsify this model.

Some critics of the SCM (e.g., Brown et al., 1999; Gleaves, 1996) have also attempted to argue that suggestive therapeutic practices can produce additional alters in patients who already meet the criteria for DID, but that these practices cannot create DID *de novo*. This assertion hinges on the assumption that iatrogenic influences can lead patients with one alter to develop additional alters, but cannot lead patients with no alters to develop one or more alters. The theoretical basis underlying this assumption has not been clearly articulated by critics of the SCM (Lilienfeld et al., 1999). Moreover, this assertion appears extremely difficult, if not impossible, to falsify given that many critics of the SCM maintain that DID alters can be "latent" (e.g., Kluft, 1992). That is, if a patient with no alters developed alters following suggestive therapeutic practices, critics of the SCM could readily maintain that this patient merely had latent alters and in fact suffered from DID all along (Piper, 1997). In addition, even some of the most vociferous proponents of the PTM acknowledge that DID can indeed be iatrogenically created in certain cases. Ross (1977), for example, estimated that approximately 17% of DID cases are predominantly iatrogenic (see also Coons, 1994). Thus, the more important question appears to be not whether DID can be created largely by iatrogenic factors, but rather what is the relative importance of iatrogenesis compared with other potential causal variables, including media influences, sociocultural factors, and individual differences in personality and psychopathology.

The Distribution of Cases of DID across Clinicians

The distribution of cases of DID across therapists is strikingly nonrandom, demonstrating that a relatively small number of clinicians account for a large number of cases of DID. For example, a 1992 survey study in

Switzerland revealed that 66% of DID diagnoses were made by 0.09% (!) of all clinicians. Moreover, 90% of respondents reported that they had never seen a single patient with DID, whereas three psychiatrists reported that they had seen over 20 patients with DID (Modestin, 1992). Ross et al. (1989) reported that members of the International Society for the Study of Multiple Personality and Dissociation (now called the ISSTD) were between 10 and 11 times more likely than members of the Canadian Psychiatric Association to report having seen a case of DID. In addition, Mai (1995) found evidence for substantial variability in the number of DID diagnoses across Canadian psychiatrists and reported that the lion's share of DID diagnoses derived from a relatively small number of psychotherapists. Boysen (2011) reported that four American research teams accounted for two-thirds of all reported published cases of childhood DID ($N = 255$). In a later review, Boysen and VanBergen (2013) found that, between 2000 and 2010, an equally remarkable two-thirds of all new cases of adult DID derived from five investigative teams.

Interestingly, these findings dovetail with those of Qin, Goodman, Bottoms, and Shaver (1998), who stated that reports of satanic ritual abuse similarly derive from a small number of psychotherapists. Reports of satanic ritual abuse are closely associated with diagnoses of DID (Mulhern, 1995).

Findings on the nonrandom distribution of DID cases are compatible with several explanations. For example, such findings could be explained by positing that patients with actual or possible DID are selectively referred to DID experts. Alternatively, perhaps certain therapists are especially adept at either detecting or eliciting the actual features of DID. Nevertheless, these findings are also consistent with the SCM and with Spanos's (1994, 1996) contention that only a handful of clinicians are diagnosing DID, producing DID symptoms in their patients, or both.

At this point, the data do not permit any adjudication among these possibilities, which are not mutually exclusive. Nevertheless, these findings provide one useful test of the SCM, because if DID diagnoses were not made disproportionately by a subset of clinicians—namely, those who are ardent proponents of the DID diagnosis—the SCM would be called into question. Longitudinal investigations examining whether patients tend to exhibit the core features of DID, especially alters, prior to or following referrals to DID specialists, would help to determine whether these findings are attributable primarily to iatrogenesis, as posited by the SCM, or to either differential referral patterns or the use of more sensitive diagnostic practices, as posted by the PTM.

Role-Playing Studies

Another source of evidence in support of the SCM derives from laboratory studies of role playing. These investigations are designed to test the

hypothesis, derived from the SCM, that cues, prompts, and suggestions from a psychotherapist can trigger participants without DID to display the overt features of this condition.

In one of these studies, Spanos, Weekes, and Bertrand (1985) provided participants with suggestions for DID (e.g., "I think perhaps there might be another part of [you] that I haven't talked to") in the context of a simulated psychiatric interview. They found that many role-playing participants, but not control participants (who were not provided with these suggestions), spontaneously adopted a different name, referred to their host personality in the third person (e.g., "He"), and exhibited striking differences between the host and alter "personalities" on psychological measures (e.g., sentence completion tests and semantic differential questionnaires). In addition, most role-playing participants, but not control participants, spontaneously reported amnesia for their alters following hypnosis. It is crucial to note that participants were not explicitly told or asked to display any of these characteristics, which are similar to those exhibited by patients with DED. These findings were essentially replicated with a similar methodology by Spanos, Weekes, Menary, and Bertrand (1996; but see Frischolz, Lipman, Braun, & Sachs, 1992). Stafford and Lynn (1998) similarly found that, given adequate situational inducements, normal participants can readily role-play a variety of life history experiences often reported among patients with DID, including reports of physical, sexual, and satanic ritual abuse.

Role-playing studies have been commonly misinterpreted by critics of the SCM. For example, Gleaves (1996) argued that "to conclude that these studies prove that DID is simply a form of role-playing is unwarranted" (p. 47). Similarly, Brown et al. (1999) contended that role-playing studies do not demonstrate that DID "can be created in the laboratory" (p. 580) and that "these role enactments are not identical with alter behavior in MPD patients, nor are they proof that a major psychiatric condition, MPD, has been created" (p. 581). But role-playing studies were not designed to reproduce the full range or subjective experience of DID symptoms, nor to create DID itself, but rather to demonstrate the ease with which subtle cues and prompts can trigger normal participants to display some of the key features of this condition. The findings of these studies (e.g., Spanos et al., 1985) provide support for the SCM because they demonstrate that (1) the behaviors and reported experiences are familiar to many members of the general population and (2) individuals without DID can be readily induced to exhibit some of the key features of DID following prompts and cues, even though these specific features were not explicitly suggested to them. Were this not the case, the SCM would not be able to account for a number of the core features of DID. Role-playing studies therefore provide corroboration for one important and potentially falsifiable precondition of the SCM, although they do not provide dispositive evidence for this model (Lilienfeld et al., 1999).

Cross-Cultural Studies

As noted earlier, the SCM posits that the overt expression of multiple identity enactments is shaped substantially by cultural and historical factors. Consistent with this presupposition is the fact that until fairly recently, DID was largely unknown outside of North America (see also Hochman & Pope, 1997, for data suggesting considerably greater acceptance of DID in North American countries compared with non-North American English-speaking countries). Indeed, between 2000 and 2010, only 18% of all reported DID cases emanated from non-Western countries (Boysen & VanBergen, 2013). For example, during the 20th century, there were only 35 reported cases of DID in Japan (Sekine, 2000; see also Takahashi, 1990). In addition, until fairly recently, DID was quite rare in England, Russia, and India (Spanos, 1996). Interestingly, the cross-cultural expression of DID appears to be different in India than in North America. In the relatively rare cases of DID reported in India, the transition between alters is almost always preceded by sleep, a phenomenon not observed in North American cases of DID. Media portrayals of DID in India similarly include periods of sleep prior to the transitions between alters (North et al., 1993).

Gleaves (1996), noting that DID has recently been diagnosed in Holland (see also Sno & Schalken, 1999) and several other European countries, used this finding to argue against the SCM. Nevertheless, this finding is difficult to interpret and does not necessarily call the SCM into question. In Holland, for example, the writings of several well-known researchers (e.g., van der Hart, 1993; van der Kolk, van der Hart, & Marmar, 1996) have resulted in substantially increased media and professional attention to DID. Recent data also point to the possibility of a relatively recent increased prevalence of DID in other countries, including Turkey, Australia, Germany, and China (Martínez-Taboas, Dorahy, Sar, Middleton, & Krüger, 2013), with Turkey accounting for 79% of all non-Western cases between 2000 and 2010 (Boysen & VanBergen, 2013). Again, however, it is unclear whether such increases reflect increases in the genuine prevalence of DID in these countries or enhanced detection or creation of DID features.

Moreover, "culturally influenced" is not equivalent to "culture bound." In other words, the fact that a condition initially limited to only a few countries subsequently spreads to other countries does not necessarily indicate that this condition is independent of cultural influence. To the contrary, the fact that the features of DID are becoming better known in certain countries would lead one to expect DID to be diagnosed with increasing frequency in these countries. The spread of DID to countries in which the characteristics of this condition are becoming more familiar constitutes one important and potentially falsifiable prediction of the SCM.

DID in Childhood

If the PTM is correct, then cases of DID should sometimes be observed in childhood, prior to extensive treatment and media exposure to the expected signs and symptoms of the condition. In a review of the literature, Boysen (2011) found at best mixed support for this possibility. As he noted, childhood DID "appears to be an extremely rare phenomenon" (p. 329). Moreover, he found that reported cases of DID in childhood have almost never been observed outside of treatment. At the same time, he reported a total of 255 cases of childhood DID in the world literature, but, as noted earlier, two-thirds of these cases originated from a very small number of research groups. Although these findings are not conclusive, they raise questions concerning the potential existence of childhood DID, a phenomenon that would be predicted by the PTM.

Summary

A variety of pieces of evidence, including commonly prescribed treatment practices of DID proponents, the clinical features of patients with DID before and after psychotherapy, the distribution of cases of DID across psychotherapists, data from role-playing studies, recent cross-cultural epidemiological data, and the extremely low prevalence of childhood DID outside of treatment, provide support for several important predictions of the SCM. In addition, these data call into question a "strong" form of the PTM (e.g., Bremner, 2010; Gleaves, 1996)—viz., a version of the PTM that essentially excludes sociocultural influence as an explanation of DID's etiology and accords virtually exclusive causal import to early trauma. These data may, however, be consistent with a "weak" form of the PTM that accords a predisposing role to early trauma but also grants a substantial causal role to sociocultural influences, including iatrogenesis (e.g., Dalenberg et al., 2012). To provide more compelling support for the PTM, proponents of this model will need to make more explicit predictions that could in principle permit this model to be falsified.

THE ETIOLOGY OF DID: THE CHILD ABUSE CONTROVERSY

As noted earlier, a linchpin of the PTM is the assumption that DID is caused largely by early trauma, particularly severe abuse, in childhood. Some authors regard DID as a form or variant of PTSD (see Bremner, 2009; Gleaves, 1996). Many authors have accepted rather uncritically the claim that severe abuse is an important precursor, if not cause, of DID. For example, Gleaves et al. (2001) concluded "there is a clear body of evidence linking DID or dissociative experiences in general with a history of childhood trauma" (p. 586; see also Carson & Butcher, 1992). In contrast, our

reading of the research literature suggests a considerably more complex and ambiguous picture, and raises important questions regarding the hypothesized association between early abuse and DID.

The Corroboration of Abuse Reports among Patients with DID

A number of investigators have reported very high prevalences of early child abuse among patients with DID (see Gleaves, 1996, p. 53). Nevertheless, in virtually none of these studies was the abuse independently corroborated (e.g., Boon & Drajer, 1993; Coons, Bowman, & Milstein, 1988; Ellason, Ross, & Fuchs, 1996; Putnam et al., 1986; Ross et al., 1989; Ross, 1990; Schultz, Braun, & Kluft, 1989; Scroppo, Drob, Weinberger, & Eagle, 1998). The absence of external corroboration in these studies is problematic in light of findings that memory is considerably more malleable, reconstructive, and vulnerable to suggestion than previously believed (Loftus, 1993, 1997; Malinowski & Lynn, 1995; Zhu, Chen, Loftus, Lin, & Dong, in press; see also Lynn, Krackow, Loftus, Locke, & Lilienfeld, Chapter 8, this volume). Moreover, memories of traumatic experiences (e.g., wartime combat) are not immune to this problem (Southwick, Morgan, Nicaolaou, & Charney, 1997), suggesting that memory malleability is not limited to artificial laboratory stimuli.

In addition, the phenomenon of "effort after meaning," whereby individuals interpret potentially ambiguous events (e.g., hitting, fondling) in accord with their implicit theories regarding the causes of their conditions, further renders some reports of relatively mild or moderate physical and sexual abuse difficult to interpret without independent corroboration (see Rind, Tromovitch, & Bauserman, 1998). Furthermore, it is difficult to exclude the possibility that the same inadvertent cues emitted by therapists that promote the creation of alters may also promote the creation of false abuse memories (Spanos, 1994), although little is known about the prevalence of suggestive therapeutic practices among DID therapists. As a consequence, it is difficult to rule out the possibility that the reported association between DID and child abuse is at least partly spurious and contaminated by therapists' methods of eliciting information.

Another potential reason for emphasizing the importance of corroboration in child abuse research on patients with DID is the recent research indicating that high scorers on the Dissociative Experiences Scale (Bernstein & Putnam, 1986), who are prone to DID and other dissociative disorders, (1) exhibit a response bias toward endorsing a large number of autobiographical events on life events questionnaires, including memories of both negative and neutral life events (Giesbrecht et al., 2008; Merckelbach, Muris, Horselenberg, & Stougie, 2000); (2) are especially likely to accept as veridical misleading statements, including those concerning autobiographical events (Ost, Fellows, & Bull, 1997); and (3) tend to be highly prone to fantasy, potentially rendering them susceptible to false memories

(e.g., Giesbrecht et al., 2008, 2010; but see Dalenberg et al., 2012, for a different perspective). It has yet to be established, however, whether these findings are directly pertinent to reports of child abuse among patients with DID. All the same, these findings raise the possibility that individuals prone to DID and related conditions may be especially likely to report life events that did not occur. This possibility warrants investigation in controlled studies.

Several investigators have, however, attempted to corroborate the retrospective abuse reports of patients with DID. For example, Coons and Milstein (1986) and Coons (1994) claimed to provide objective documentation for the abuse reports of a number of patients with DID. Close inspection of these studies, however, reveals various methodological shortcomings. In neither study were diagnoses of DID made blindly of previous abuse reports. This methodological shortcoming is problematic because certain therapists may be especially likely to attempt to elicit features of DID among patients with a history of severe abuse. In the Coons (1994) study, diagnoses of DID were made only after medical histories and psychiatric records (many of which may have contained information regarding abuse histories) were reviewed. Moreover, because standardized interviews were not administered in Coons and Milstein (1986) and were administered only to an unknown number of participants in Coons (1994), the possibility of diagnostic bias is heightened. Finally, the patients in Coons (1994) "were diagnosed personally by the first author over an 11 year period" (p. 106). Because there is no evidence concerning whether these patients met the criteria for DID prior to treatment, the possibility of iatrogenic influence is difficult to exclude.

Lewis et al. (1997) reported findings from a study of 12 murderers with DID that, in their words, "establishes, once and for all, the linkage between early severe child abuse and dissociative identity disorder" (p. 1703). Some authors have cited Lewis et al.'s findings as providing strong evidence for the corroboration of abuse reports among patients with DID (e.g., Gleaves et al., 2001). Nevertheless, Lewis et al.'s objective documentation of abuse was often quite vague (see also Klein, 1999). For example, in several cases, there are indications only that the "mother [was] charged as unfit" or that "emergency room records report[ed] severe headaches"). In addition, their findings are difficult to interpret for several other reasons. First, the objective documentation of childhood DID symptoms was similarly vague in many cases and was often based on the presence of imaginary playmates and other features (e.g., marked mood changes) that are extremely common in childhood. Second, because violent individuals tend to have high rates of abuse in childhood (Widom, 1989), Lewis et al.'s findings are potentially attributable to the confounding of DID with violence. Third, diagnoses of DID were not performed blindly with respect to knowledge of reported abuse history. Fourth, the murderers' handwriting samples, which differed over time and were used by Lewis et al. to buttress the claim that these

individuals had DID, were not systematically evaluated by graphoanalysts or compared with the handwriting samples of normals over time. Fifth, the possibility of malingering (which is often a particular problem among criminals) was not systematically evaluated with psychometric indices. These methodological limitations raise serious questions regarding Lewis et al.'s claim that their study provides definitive evidence of an association between early child abuse and later DID.

A more indirect approach to the corroboration of child abuse among patients with DID was adopted in a widely publicized study by Tsai, Condie, Wu, and Chang (1999), who used magnetic resonance imaging with a 47-year-old female with DID. Reasoning from previous investigations that had reported a reduction in hippocampal volume following combat trauma (e.g., Bremner, Randall, Scott, & Bronen, 1995) and early child abuse (Bremner, Randall, Vermetten, & Staib, 1997; Stein et al., 1997), Tsai et al. hypothesized that patients with DID (given their presumed history of early abuse) would similarly exhibit decreased hippocampal volume. As predicted, they found significant bilateral reductions in hippocampal volume in these patients, which is broadly consistent with predictions derived from the PTM. Nevertheless, this finding must be interpreted cautiously for two major reasons. First, because it is based on only one patient, its generalizability to other individuals with DID is unclear. Second, decreased hippocampal volume is not specific to PTSD or to other conditions secondary to trauma; it has also been reported in schizophrenia (Nelson, Saykin, Flashman, & Riordan, 1998) and depression (Bremner et al., 2000). Consequently, decreased hippocampal volume may be a nonspecific marker of long-term stress (Sapolsky, 2000), which is present in many psychiatric conditions.

Moreover, several pieces of data raise questions regarding the veracity of some reports of child abuse in studies of DID, and underscore the importance of corroborating these reports. In the study by Ross et al. (1989), 26% of patients with DID reported being abused prior to age 3, and 10.6% reported being abused prior to age 1. Similarly, Dell and Eisenhower (1990) noted that 4 of 11 adolescent patients with DID reported that their first alter emerged at age 2 or earlier, and 2 of these patients reported that their first alter emerged between the ages 1 of 2. Memories reported prior to age 3 are of extremely questionable validity, and it is almost universally accepted that adults and adolescents are unable to remember events that occurred prior to age 1 (Fivush & Hudson, 1990). It is possible that the memories reported in these studies were accurate, but that they were dated incorrectly. Nonetheless, the nontrivial percentages of individuals in Ross et al. (1991) and Dell and Eisenhower (1990) who reported abuse and the emergence of alters at very young ages raise concerns regarding the accuracy of these memories.

Finally, Ross and Norton (1989) found that patients with DID who had been hypnotized reported significantly higher rates of sexual and physical

abuse than patients with DID who had not been hypnotized. Because there is little evidence that hypnosis enhances the accuracy of memory (Lynn, Lock, Myers, & Payne, 1997), this finding is consistent with the possibility that hypnosis produces an increased rate of false abuse reports. Nevertheless, this conclusion must remain tentative in view of the absence of independent corroboration of the abuse reports and the correlational nature of Ross and Norton's data.

Interpretation of the Child Abuse–DID Association

Even if the child abuse reports of most patients with DID were corroborated, several important questions arise concerning the interpretation of these reports. In particular, it remains to be determined whether a history of child abuse is (1) more common among patients with DID than among psychiatric patients in general and (2) causally associated with risk for subsequent DID.

With respect to the first issue, base rates and referral biases pose potential problems when interpreting the child abuse data. Because the prevalence of reported child abuse among psychiatric patients in general tends to be high (e.g., Pope & Hudson, 1992), these data are difficult to interpret without a psychiatric comparison group. This omission is particularly concerning in view of findings that DID overlaps substantially with a host of psychiatric conditions (e.g., borderline personality disorder, bipolar disorder) that may sometimes be associated with elevated rates of child abuse (Ross & Ness, 2010).

Moreover, the co-occurrence between reported abuse and DID could be a consequence of several selection artifacts that increase the probability that individuals with multiple problems seek treatment. Berksonian bias (Berkson, 1946) is a mathematical artifact that results from the fact that an individual with two problems can seek treatment for either problem. Clinical selection bias (see duFort, Newman, & Bland, 1993) reflects the increased likelihood that patients with one problem will seek treatment if they subsequently develop another problem. Either or both of these artifacts could lead to the apparent relation between child abuse and DID. Indeed, Ross (1991) found that nonclinical participants with DID reported substantially lower rates of child abuse than did patients with DID recruited from a clinical population. This finding is consistent with the hypothesis that selection biases account at least partly for the high levels of co-occurrence between reported child abuse and DID. Moreover, Ross et al. (1989) reported that American psychiatrists reported a substantially higher prevalence of child abuse among patients with DID (81.2%) than did Canadian psychiatrists (45.5%). This finding suggests the possibility of biases in the assessment or elicitation of child abuse reports and raises questions concerning the claim that child abuse is necessary for most cases of DID (Spanos, 1994).

If a clear correlation between early child abuse and DID could be

unambiguously demonstrated, it would still be necessary to demonstrate that this abuse plays a causal role in subsequent DID. This task will be difficult given the fact that studies of early abuse in patients with DID are necessarily quasi-experimental. Nevertheless, data from causal modeling studies could help to shed light on this question. In addition, studies of monozygotic (identical) twins discordant for early abuse history could help to provide more compelling evidence for a causal role of abuse in DID. Specifically, if it could be demonstrated that only the MZ twin with a history of early abuse exhibited significant levels of dissociative features (including features of DID), then this finding would buttress the contention that early abuse, rather than a host of other potential nuisance variables that distinguish dissociative patients from other individuals (e.g., genetic differences in the propensity toward suggestibility), plays an etiological role in DID.

Summary

The PTM hinges on the assumption that early trauma, particularly child abuse, is a precursor of, and risk factor for, DID. Consistent with this assumption, many authors have found that a large proportion, and probably a majority, of patients with DID report a history of early and sometimes severe child abuse. Nevertheless, careful inspection of this literature raises significant questions concerning the child abuse–DID link. Most of the reported confirmations of this association derive from studies lacking objective corroboration of child abuse (e.g., Ross et al., 1990). Moreover, even those studies that purport to provide such corroboration (e.g., Coons, 1994; Lewis et al., 1997) are plagued by numerous methodological shortcomings. In addition, the reported high levels of child abuse among patients with DID may be attributable to selection and referral biases common in psychiatric samples, as well as high levels of comorbidity between DID and other conditions. Finally, it is unclear whether early abuse plays a causal role in DID. These methodological limitations do not exclude a potential etiological role for early trauma in DID, but they suggest the need for further controlled research before strong conclusions regarding the child abuse–DID link (e.g., Gleaves, 1996; Gleaves et al., 2001) can be drawn.

CONCLUSIONS

The literature on DID has recently been engulfed in numerous divisive controversies (see also Elzinga et al., 1998). In particular, there has been substantial scientific disagreement over whether DID is (1) a "genuine" condition, (2) truly characterized by the coexistence of multiple indwelling and fully developed personalities, (3) a socially constructed product of iatrogenic, media, and cultural influences, and (4) a consequence of

early childhood trauma, particularly child abuse. As we have argued, controversy (1) is actually a pseudoissue, as there is no longer much dispute that DID "exists" in the sense that most individuals with this condition genuinely exhibit signs and symptoms of psychopathology and experience intense subjective distress. We therefore urge that authors no longer frame the DID debate in terms of this condition's "existence" (e.g., Gleaves, 1996).

Controversy (2) is difficult or impossible to resolve with existing data, although it is clear that most DID researchers, even those who are fervent proponents of the PTM (e.g., Ross, 1997), do not believe that the alters of patients with DID constitute fully developed and independent personalities (e.g., American Psychiatric Association, 2013). Moreover, the proposition that alters constitute fully developed and independent personalities poses significant challenges to models of DID's etiology and development, particularly for patients with very large numbers of alters.

Perhaps the primary controversy surrounding DID is the question of whether DID is a socially constructed and culturally influenced condition rather than a naturally occurring response to early trauma. As we have argued elsewhere (Lilienfeld et al., 1999), a number of important lines of evidence converge to provide support for the SCM. In particular, 11 findings are consistent with the major theses of the SCM:

1. The number of patients with DID has increased dramatically over the past few decades (Elzinga et al., 1998).
2. The number of alters per DID individual has similarly increased over the past few decades (North et al., 1993), although the number of alters at the time of initial diagnosis appears to have remained constant (Ross et al., 1989).
3. Both of these increases coincide with dramatically increased therapist and public awareness of the major features of DID (Fahy, 1988).
4. Treatment techniques for DID advocated by some proponents of the PTM (e.g., ISSTD, 2011) may reinforce patients' displays of multiplicity (Phelps, 2000), reify alters as distinct personalities, and encourage patients to establish contact with presumed latent alters (Spanos, 1994, 1996).
5. Many or most patients with DID show few or no clear-cut signs of this condition (e.g., alters) prior to psychotherapy (Kluft, 1984).
6. The number of alters per DID individual tends to increase substantially over the course of DID-oriented psychotherapy (Piper, 1997).
7. Psychotherapists who use hypnosis tend to have more patients with DID in their caseloads than do psychotherapists who do not use hypnosis (Powell & Gee, 1999).
8. The majority of diagnoses of DID derive from a relatively small

number of psychotherapists, many of whom are specialists in DID (Boysen, 2011; Boysen & VanBergen, 2013; Mai, 1995).

9. Laboratory studies suggest that nonclinical participants who are provided with appropriate cues and prompts can reproduce many of the overt features of DID (Spanos et al., 1985).
10. Until fairly recently, diagnoses of DID were limited largely to North America, where the condition has received widespread media publicity (Spanos, 1996), although DID is now being diagnosed with considerable frequency in some countries (e.g., Holland, Turkey) in which it has recently become more widely publicized (Boysen & VanBergen, 2013).
11. Childhood DID appears to be extremely rare or nonexistent outside of treatment (Boysen, 2011).

These 11 sources of evidence do not imply, however, that DID can typically be created *in vacuo* by iatrogenic or sociocultural influences. As noted earlier, a large proportion of patients with DID have histories of co-occurring psychopathology, particularly borderline personality disorder and bipolar disorder (Ganaway, 1995; Lynn et al., 2011). Moreover, the SCM is entirely consistent with the possibility that familial factors, such as poor attachment or neglectful parenting, or genetic vulnerabilities toward emotional dysregulation, may increase risk for DID-like symptoms. Therefore, it seems plausible that iatrogenic and sociocultural influences often operate on a backdrop of preexisting psychopathology, life stressors, and genetic influences, and exert their impact primarily on individuals who are seeking a causal explanation for their instability, identity problems, and impulsive and seemingly inexplicable behaviors.

We should also note that several of these 11 sources of evidence are fallible and open to multiple causal interpretations (Lilienfeld et al., 1999). For example, the finding that the number of alters per individual tends to increase over the course of psychotherapy is potentially consistent with the assertion (Ross, 1997) that psychotherapy for DID is often accompanied by a progressive uncovering of previously latent alters. In addition, the finding that diagnoses of DID have increased dramatically over the past few decades is potentially attributable to the advent of superior diagnostic and assessment practices among DID practitioners. Moreover, as noted earlier, diagnoses of several other psychiatric disorders, including PTSD and obsessive compulsive disorder (OCD), have increased over the past three decades (Zohar, 1998).

Although none of these 11 lines of evidence is by itself dispositive, the convergence of evidence across all of these sources of data provides a potent argument for the validity of the SCM (Lynn et al., 2014; see also Lynn & Pintar, 1997). Our conclusions differ sharply from those of Brown et al. (1999), who contended that "the entire data base of 'scientific evidence' [for the SCM] consists of a grand total of three experimental studies—all

coming out of the same laboratory" (p. 617). Brown et al. were referring to the laboratory role-playing studies of Spanos and his colleagues (e.g., Spanos et al., 1985).

Nevertheless, Brown et al. (1999) drew this conclusion only because they restricted themselves entirely to strictly experimental studies (i.e., those involving random assignment to conditions and manipulation of a discrete independent variable) when evaluating the scientific status of the SCM. This approach is grossly underinclusive because a variety of lines of quasi-experimental and observational evidence (e.g., the higher rates of psychopathology of patients with DID after versus before psychotherapy, the markedly nonrandom distribution of DID cases of DID across practitioners) are directly relevant to the validity of the SCM. In many sophisticated "hard" sciences, including geology, astronomy, meteorology, and paleontology, nonexperimental evidence is used routinely to test causal hypotheses (although this evidence can rarely, if ever, be used to prove causal relationships), and the same evidentiary guidelines should hold in psychology. Indeed, as 19th-century philosopher William Whewell observed, most scientific hypotheses are tested by evaluating the "consilience of evidence" across diverse and maximally independent sources of information (Shermer, 2001). The consilience of evidence for the SCM is striking and strongly suggests that iatrogenic and sociocultural influences play at least some etiological role in DID.

This conclusion does not imply, however, that the PTM has been falsified or should be abandoned. With respect to the fourth major controversy examined in this chapter, namely, the child abuse–DID link, extant studies provide relatively weak support for the contention that child abuse is a precursor or potent causal risk factor for DID (cf. Dalenberg et al., 2012; Gleaves et al., 2001). Nevertheless, this possibility cannot be excluded on the basis of existing evidence. Studies that provide corroborated abuse reports and psychiatric comparison groups, and that control for selection and referral biases, are required to bring clarity to this methodologically complex area (Lilienfeld et al., 1999). In addition, causal modeling studies may help to exclude alternative hypotheses for the high levels of co-occurrence between reported child abuse and later DID. If such abuse can be corroborated and shown to be associated with risk for later DID, such studies will be especially informative if they incorporate potential third variables that could account for this correlation (e.g., adverse home environment).

If future studies provide more convincing evidence for the child abuse–DID association, such evidence might necessitate a rapprochement between the SCM and PTM. Indeed, some important aspects of these two models may ultimately prove commensurable. For example, trauma might predispose individuals to develop high levels of fantasy-proneness (Giesbrecht et al., 2008; Lynn, Rhue, & Green, 1988), absorption (Tellegen & Atkinson, 1974), and/or sleep problems, all of which may predispose toward dissociation (van der Kloet, Giesbrecht, Lynn, Merckelbach, & de Zutter, 2012). In turn, these psychological characteristics may render individuals susceptible

to the kinds of iatrogenic and cultural influences posited by the SCM, thereby increasing the likelihood that they will develop DID and related dissociative disorders following exposure to suggestive influences. This and even more sophisticated etiological models of DID have yet to be subjected to direct empirical tests.

Given the converging support for the SCM across multiple sources of evidence, however, we believe that the burden of proof now falls squarely on proponents of the PTM to provide more compelling evidence for this position (cf. Bremner, 2010; Brown et al., 1999; Dalenberg et al., 2012). If they are successful, the multiple controversies that have swirled around the diagnosis of DID virtually since its inception could prove closer to a satisfactory resolution.

GLOSSARY

Alter: One of the presumed "personalities" or "personality states" of individuals with dissociative identity disorder.

Borderline personality disorder: A personality disorder characterized by identity confusion, mood instability, erratic and unpredictable interpersonal relationships, and impulsive and self-damaging behaviors, among other features.

Dissociation: A defense mechanism ostensibly characterized by the compartmentalization or "walling off" of negative experiences from consciousness.

Dissociative disorders: A set of disorders, including dissociative identity disorder, characterized by disturbances in memory, identity, consciousness, and/or perception of the external environment.

Dissociative identity disorder (DID): A condition, known formerly as multiple personality disorder, characterized by the presence of distinct personalities or personality states that recurrently take control over the individual's behavior. This condition is also characterized by marked memory gaps for autobiographical information.

Host personality: The "original" or primary personality of the individual with DID.

Iatrogenic: Adverse effect produced by physicians or mental health professionals.

Inner self-helper: As proposed by Allison (1974), a part of the personality of individuals with DID that is aware of what is occurring to the alters and can assist in their integration.

Posttraumatic model: A model positing that DID is a naturally occurring response to childhood trauma, particularly child physical and/or sexual abuse.

Sociocognitive model: A model positing that dissociative identity disorder is a socially constructed condition resulting primarily from inadvertent therapist prompting, media influences, and sociocultural expectations regarding the presumed features of this condition.

REFERENCES

Acocella, J. (1998, April 6). The politics of hysteria. *New Yorker*, pp. 64–79.

Aldridge-Morris, R. (1989). *Multiple personality: An exercise in deception*. Hillsdale, NJ: Erlbaum.

Allen, J. J. B., & Iacono, W. G. (2001). Assessing the validity of amnesia in

dissociative identity disorder: A dilemma for the DSM and the courts. *Psychology, Public Policy, and Law, 7*, 311–344.

Allen, J. J. B., & Movius, H. L. (2000). The objective assessment of amnesia in dissociative identity disorder using event-related potentials. *International Journal of Psychophysiology*, *38*, 21–41.

Allison, R. (1974). A new treatment approach for multiple personality. *American Journal of Clinical Hypnosis*, *17*, 15–32.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Arrigo, J. M., & Pezdek, K. (1998). Textbook models of multiple personality: Source, bias, and social consequences. In S. J. Lynn (Ed.), *Truth in memory* (pp. 372–393). New York: Guilford Press.

Bahnson, C. B., & Smith, K. (1975). Autonomic changes in a multiple personality. *Psychosomatic Medicine, 37*, 85–86.

Barber, T. X. (1979). Suggested ("hypnotic") behavior: The trance paradigm versus an alternative paradigm. In E. Fromm & R. E. Shor (Eds.), *Hypnosis: Developments in research and new perspectives* (2nd ed., pp. 217–271). New York: Aldine.

Belli, R. F., Winkielman, P., Read, J. D., Schwarz, N., & Lynn, S. J. (1998). Recalling more childhood events leads to judgments of poorer memory: Implications for the recovered/false memory debate. *Psychonomic Bulletin and Review*, *5*, 318–323.

Berkson, J. (1946). Limitations of the application of the four-fold table analysis to hospital data. *Biometrics Bulletin*, *2*, 47–53.

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease*, *174*, 727–735.

Bliss, E. L. (1980). Multiple personalities: A report of 14 cases with implications for schizophrenia and hysteria. *Archives of General Psychiatry*, *37*, 1388–1397.

Bloch, J. P. (1991). *Assessment and treatment of multiple personality and dissociative disorders*. Sarasota, FL: Professional Resource Press.

Boon, S., & Draijer, N. (1993). Multiple personality disorder in the Netherlands: A clinical investigation of 71 cases. *American Journal of Psychiatry*, *150*, 489–494.

Boor, M. (1982). The multiple personality epidemic: Additional cases and inferences regarding diagnosis, etiology, dynamics, and treatment. *Journal of Nervous and Mental Disease*, *170*, 302–304.

Bottoms, B. L., Shaver, P. R., & Goodman, G. S. (1996). An analysis of ritualistic and religion-related child abuse allegations. *Law and Human Behavior*, *20*, 1–34.

Boysen, G. A. (2011). The scientific status of childhood dissociative identity disorder: A review of published research. *Psychotherapy and Psychosomatics*, *80*, 329–334.

Boysen, G. A., & VanBergen, A. (2013). A review of published research on adult dissociative identity disorder: 2000–2010. *Journal of Nervous and Mental Disease*, *201*, 5–11.

Brand, B. L., Classen, C. C., Lanius, R., Loewenstein, R. J., McNary, S. W., Pain, C., et al. (2009). A naturalistic study of dissociative identity disorder and dissociative disorder not otherwise specified patients treated by community clinicians. *Psychological Trauma: Theory, Research, Practice, and Policy*, *1*, 153–171.

Brand, B. L., Classen, C. C., McNary, S. W., & Zaveri, P. (2009). A review of dissociative disorders treatment studies. *The Journal of Nervous and Mental Disease*, *197*, 646–654.

Brand, B. L., McNary, S. W., Myrick, A. C., Classen, C. C., Lanius, R., Loewenstein, R. J., et al. (2012). A longitudinal naturalistic study of patients with dissociative disorders treated by community clinicians. *Psychological Trauma: Theory, Research, and Practice*, *5*, 301–308.

Braun, B. G. (1980). Hypnosis for multiple personalities. In H. J. Wain (Ed.), *Clinical hypnosis in medicine* (pp. 209–217). Chicago: Year Book Medical.

Braun, B. G. (1986). Issues in the psychotherapy of multiple personality disorder. In B. G. Braun (Ed.), *Treatment of multiple personality disorder* (pp. 1–28). Washington, DC: American Psychiatric Press.

Bremner, J. D. (1999). Does stress damage the brain? *Biological Psychiatry*, *45*, 797–805.

Bremner, J. D. (2010). Cognitive processes in dissociation: Comment on Giesbrecht et al. (2008). *Psychological Bulletin*, *136*, 1–11.

Bremner, J. D., Narayan, M., Anderson, E. R., Staib, L. H., Miller, H. L., & Charney, D. S. (2000). Hippocampal volume reduction in major depression. *American Journal of Psychiatry*, *157*, 115–117.

Bremner, J. D., Randall, P., Scott, T. M., & Bronen, R. (1995). MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. *American Journal of Psychiatry*, *152*, 973–981.

Bremner, J. D., Randall, P., Vermetten, E., & Staib, L. (1997). Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse: A preliminary report. *Biological Psychiatry*, *41*, 23–32.

Brende, J. O. (1984). The psychophysiologic manifestations of dissociation: Electodermal responses in a multiple personality patient. *Psychiatric Clinics of North America*, *7*, 41–50.

Brown, D., Frischoltz, E. J., & Scheflin, A. W. (1999). Iatrogenic dissociative identity disorder: An evaluation of the scientific evidence. *Journal of Psychiatry and Law*, *27*, 549–637.

Byrne, P. (2001). The butler(s) DID it—dissociative identity disorder in cinema. *Medical Humanities*, *27*, 26–29.

Carson, R. C., & Butcher, J. N. (1992). *Abnormal psychology and modern life* (9th ed.). New York: HarperCollins.

Choe, B. M., & Kluft, R. P. (1995). The use of the DES in studying treatment outcome with dissociative identity disorder: A pilot study. *Dissociation: Progress in the Dissociative Disorders*, *8*, 160–164.

Coons, P. M. (1984). The differential diagnosis of multiple personality disorder: A comprehensive review. *Psychiatric Clinics of North America*, *7*, 51–67.

Coons, P. M. (1994). Confirmation of childhood abuse in child and adolescent cases of multiple personality disorder and dissociative identity disorder not otherwise specified. *Journal of Nervous and Mental Disease*, *182*, 461–464.

Coons, P. M., Bowman, E. S., & Milstein, V. (1988). Multiple personality disorder: A clinical investigation of 50 cases. *Journal of Nervous and Mental Disease*, *176*, 519–527.

Coons, P. M., & Milstein, V. (1986). Psychosexual disturbances in multiple personality: Characteristics, etiology, and treatment. *Journal of Clinical Psychiatry*, *47*, 106–111.

Cormier, J. F., & Thelen, M. H. (1998). Professional skepticism of multiple personality disorder. *Professional Psychology: Research and Practice*, *29*, 163–167.

Dalenberg, C. J., Brand, B. L., Gleaves, D. H., Dorahy, M. J., Loewenstein, R. J., Cardeña, E., et al. (2012). Evaluation of the evidence for the trauma and fantasy models of dissociation. *Psychological Bulletin*, *138*, 550–588.

Dell, P. F. (2001). Why the diagnostic criteria for dissociative identity disorder should be changed. *Journal of Trauma and Dissociation*, *2*, 7–37.

Dell, P. F. (2006). A new model of dissociative identity disorder. *Psychiatric Clinics of North America*, *29*, 1–26.

Dell, P. F., & Eisenhower, J. W. (1990). Adolescent multiple personality disorder: A preliminary study of eleven cases. *Journal of the American Academy of Child and Adolescent Psychiatry*, *29*, 359–366.

Dell, P. F., & O'Neil, J. A. (Eds.). (2010). *Dissociation and the dissociative disorders: DSM-V and beyond*. New York: Routledge.

Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown.

Dorahy, M. J. (2001). Dissociative identity disorder and memory dysfunction: The current state of experimental research and its future directions. *Clinical Psychology Review*, *21*, 771–795.

duFort, G. G., Newman, S. C., & Bland, R. C. (1993). Psychiatric comorbidity and treatment seeking: Sources of selection bias in the study of clinical populations. *Journal of Nervous and Mental Disease*, *181*, 467–474.

Dunn, G. E., Paolo, A. M., Ryan, J. J., & van Fleet, J. N. (1994). Belief in the existence of multiple personality disorder among psychologists and psychiatrists. *Journal of Clinical Psychology*, *50*, 454–457.

Ellason, J. W., & Ross, C. A. (1997). Two-year follow-up of inpatients with dissociative identity disorder. *American Journal of Psychiatry*, *154*, 832–839.

Ellason, J. W., Ross, C. A., & Fuchs, D. L. (1996). Lifetime Axis I and Axis II comorbidity and childhood trauma history in dissociative identity disorder. *Psychiatry: Interpersonal and Biological Processes*, *59*, 255–266.

Elzinga, B. M., van Dyck, R., & Spinhoven, P. (1998). Three controversies about dissociative identity disorder. *Clinical Psychology and Psychotherapy*, *5*, 13–23.

Fahy, T. A. (1988). The diagnosis of multiple personality disorder: A critical review. *British Journal of Psychiatry*, *153*, 597–606.

Farrell, H. M. (2011). Dissociative identity disorder: Medicolegal challenges. *Journal of the American Academy of Psychiatry and the Law Online*, *39*, 402–406.

Fivush, R., & Hudson, J. A. (Eds.). (1990). *Knowing and remembering in young children*. New York: Cambridge University Press.

Frances, A., & First, M. B. (1998). *Your mental health: A layman's guide to the psychiatrist's Bible*. New York: Scribner.

Frischholz, E. J., Lipman, L. S., Braun, B. G., & Sachs, R. (1992). Suggested posthypnotic amnesia in psychiatric patients and normals. *American Journal of Clinical Hypnosis*, *35*, 29–39.

Ganaway, G. K. (1989). Historical versus narrative truth: Clarifying the role of exogenous trauma in the etiology of MPD and its variants. *Dissociation*, *2*, 205–220.

Ganaway, G. K. (1995). Hypnosis, childhood trauma, and dissociative identity disorder: Toward an integrative theory. *International Journal of Clinical and Experimental Hypnosis*, *43*, 127–144.

Gantt, L., & Tinnin, L. W. (2007). Intensive trauma therapy of PTSD and dissociation: An outcome study. *Arts in Psychotherapy*, *34*, 69–80.

Giesbrecht, T., Lynn, S. J., Lilienfeld, S. O., & Merckelbach, H. (2008). Cognitive processes in dissociation: An analysis of core theoretical assumptions. *Psychological Bulletin*, *134*, 617–647.

Giesbrecht, T., Lynn, S. J., Lilienfeld, S. O., & Merckelbach, H. (2010). Cognitive processes, trauma, and dissociation—Misconceptions and misrepresentations: Reply to Bremner (2010). *Psychological Bulletin*, *136*, 7–11.

Gleaves, D. H. (1996). The sociocognitive model of dissociative identity disorder: A reexamination of the evidence. *Psychological Bulletin*, *120*, 42–59.

Gleaves, D. H., May, M. C., & Cardena, E. (2001). An examination of the diagnostic validity of dissociative identity disorder. *Clinical Psychology Review*, *21*, 577–608.

Goulding, R. A., & Schwartz, R. C. (2002). *The mosaic mind: Empowering the tormented selves of child abuse survivors*. Oak Park, IL: Trailheads.

Hacking, I. (1995). *Rewriting the soul: Multiple personality and the science of memory*. Princeton, NJ: Princeton University Press.

Hochman, J., & Pope, H. G. (1997). Debating dissociative diagnoses. *American Journal of Psychiatry*, *153*, 887–888.

Huntjens, R. J., Verschuere, B., & McNally, R. J. (2012). Inter-identity autobiographical amnesia in patients with dissociative identity disorder. *PloS One*, *7*(7), e40580.

Hyler, S. E., & Spitzer, R. L. (1978). Hysteria split asunder. *American Journal of Psychiatry*, *135*, 1500–1504.

Infinite Mind, The. (1998, June 8). Radio Broadcast, National Public Radio.

International Society for the Study of Trauma and Dissociation (ISSTD). (2011). Guidelines for treating dissociative identity disorder in adults, third revision. *Journal of Trauma and Dissociation*, *12*, 115–187.

Janet, P. (1927). *La pensée intérieure et ses troubles*. Paris: Chahine.

Jang, K. L., Paris, J., Zweig-Frank, H., & Livesley, W. J. (1998). Twin study of dissociative experience. *Journal of Nervous and Mental Disease*, *186*, 345–351.

Klein, D. F. (1999). Multiples: No amnesia for child abuse. *American Journal of Psychiatry*, *156*, 976–977.

Kluft, R. P. (1984). Treatment of multiple personality disorder: A study of 33 cases. *Psychiatric Clinics of North America, 7*, 20–29.

Kluft, R. P. (1991). Multiple personality disorder. In A. Tasman & S. M. Goldfinger (Eds.), *American Psychiatric Press review of psychiatry* (Vol. 10, pp. 161–188). Washington, DC: American Psychiatric Association Press.

Kluft, R. P. (1993). Multiple personality disorders. In D. Spiegel (Ed.), *Dissociative disorders: A clinical review* (pp. 17–44). Lutherville, MD: Sidran Press.

Lewis, D. O., Yeager, C. A., Swica, Y., Pincus, J. H., & Lewis, M. (1997). Objective documentation of child abuse and dissociation in 12 murderers with dissociative identity disorder. *American Journal of Psychiatry*, *143*, 1703–1710.

Lilienfeld, S. O., Lynn, S. J., Kirsch, I., Chaves, J. F., Sarbin, T. R., Ganaway, G. K., et al. (1999). Dissociative identity disorder and the sociocognitive model: Recalling the lessons of the past. *Psychological Bulletin*, *125*, 507–523.

Loftus, E. R. (1993). The reality of repressed memories. *American Psychologist*, *48*, 518–537.

Loftus, E. R. (1997, September). Creating false memories. *Scientific American*, pp. 70–75.

Ludwig, A. M., Brandsma, J. M., Wilbur, C. B., Bendefeldt, F., & Jameson, D. H. (1972). The objective study of a multiple personality: Or, Are four heads better than one? *Archives of General Psychiatry*, *26*, 298–310.

Lynn, S. J., Berg, J., Lilienfeld, S. O., Merckelbach, H., Giesbrecht, T., Accardi, M., et al. (2012a). Dissociative disorders. In M. Hersen & D. C. Beidel (Eds.), *Adult psychopathology and diagnosis* (pp. 497–538). New York: Wiley.

Lynn, S. J., Lilienfeld, S. O., Merckelbach, H., Giesbrecht, T., & van der Kloet, D. (2012b). Dissociation and dissociative disorders: Challenging conventional wisdom. *Current Directions in Psychological Science*, *21*, 48–53.

Lynn, S. J., Lilienfeld, S. O., Merckelbach, H., Giesbrecht, T., McNally, R. J., Loftus, E. F., et al. (2014). The trauma model of dissociation: Inconvenient truths and stubborn fictions. *Comment on Dalenberg et al. (2012)*, *140*, 896–910.

Lynn, S. J., Lock, T. G., Myers, B., & Payne, D. (1997). Recalling the unrecallable: Should hypnosis be used to recover memories in psychotherapy? *Current Directions in Psychological Science*, *6*, 79–83.

Lynn, S. J., & Pintar, J. (1997). A social narrative model of dissociative identity disorder. *Australian Journal of Clinical and Experimental Hypnosis*, *25*, 1–7.

Lynn, S. J., Rhue, J. W., & Green, J. P. (1988). Multiple personality and fantasy proneness: Is there an association or dissociation. *British Journal of Experimental and Clinical Hypnosis*, *5*, 138–142.

Mai, F. M. (1995). Psychiatrists' attitudes to multiple personality disorder: A questionnaire study. *Canadian Journal of Psychiatry*, *40*, 154–157.

Malinowski, P., & Lynn, S. J. (1995, August). *The pliability of early memory reports*. Paper presented at the annual convention of the American Psychological Association, Washington, DC.

Marmer, S. S. (1998, December). Should dissociative identity disorder be considered a bona fide psychiatric diagnosis? Retrieved from *www.clinicalpsychiatrynews.com*.

Martínez-Taboas, A., Dorahy, M., Sar, V., Middleton, W., & Krüger, C. (2013). Growing not dwindling: International research on the worldwide phenomenon of dissociative disorders. *Journal of Nervous and Mental Disease*, *201*, 353–354.

McHugh, P. R. (1993). Multiple personality disorder. *Harvard Mental Health Newsletter*, *10*(3), 4–6.

McHugh, P. R. (2008). *Try to remember: Psychiatry's clash over meaning, memory, and mind*. New York: Dana Press.

Merckelbach, H., Devilly, G. J., & Rassin, E. (2002). Alters in dissociative identity disorder: Metaphors or genuine entities? *Clinical Psychology Review*, *22*, 481–497.

Merckelbach, H., Muris, P., Horselenberg, R., & Stougie, S. (2000). Dissociative experiences, response bias, and fantasy proneness in college students. *Personality and Individual Differences*, *28*, 49–57.

Merskey, H. (1992). The manufacture of personalities: The production of multiple personality disorder. *British Journal of Psychiatry*, *160*, 327–340.
Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law*, *6*, 1–16.
Modestin, J. (1992). Multiple personality disorder in Switzerland. *American Journal of Psychiatry*, *149*(1), 88–92.
Mulhern, S. (1995). Les aleas de la therapie des reminscences: Le trouble de la personalite. In M. Gabel, S. Lebovic, & P. Mazet (Eds.), *Le traumatisme de l'inceste*. Paris: Presses Universitaires de France.
Nathan, D. (2011). *Sybil exposed*. New York: Free Press
Nelson, M. D., Saykin, A. J., Flashman, L. A., & Riordan, H. J. (1988). Hippocampal volume reduction as assessed by magnetic resonance imaging: A meta-analytic study. *Archives of General Psychiatry*, *55*, 433–440.
North, C. S., Ryall, J.-E. M., Ricci, D. A., & Wetzel, R. D. (1993). *Multiple personalities, multiple disorders*. New York: Oxford University Press.
Orne, M. T., Dinges, D. F., & Orne, E. C. (1984). On the differential diagnosis of multiple personality in the forensic context. *International Journal of Clinical and Experimental Hypnosis*, *32*, 118–169.
Ost, J., Fellows, B., & Bull, R. (1997). Individual differences and the suggestibility of human memory. *Contemporary Hypnosis*, *14*, 132–137.
Paris, J. (2012). The rise and fall of dissociative identity disorder. *Journal of Nervous and Mental Disease*, *200*, 1076–1079.
Phelps, B. J. (2000). Dissociative identity disorder: The relevance of behavior analysis. *Psychological Record*, *50*, 235–249.
Piper, A. (1993). "Truth serum" and "recovered memories" of sexual abuse: A review of the evidence. *Journal of Psychiatry and Law*, *21*, 447–471.
Piper, A. (1997). *Hoax and reality: The bizarre world of multiple personality disorder*. Northvale, NJ: Aronson.
Piper, A., & Merskey, H. (2004). The persistence of folly: A critical examination of dissociative identity disorder: Part I. The excesses of an improbable concept. *Canadian Journal of Psychiatry*, *49*, 592–600.
Pope, H. G., Jr., Barry, S., Bodkin, A., & Hudson, J. I. (2005). Tracking scientific interest in the dissociative disorders: A study of scientific publication output 1984–2003. *Psychotherapy and Psychosomatics*, *75*(1), 19–24.
Pope, H. G., & Hudson, J. I. (1992). Is childhood sexual abuse a risk factor for bulimia nervosa? *American Journal of Psychiatry*, *149*, 455–463.
Pope, H. G., Oliva, P. S., Hudson, J. I., Bodkin, J. A., & Gruber, A. J. (1999). Attitudes toward DSM-IV dissociative disorders diagnoses among board-certified American psychiatrists. *American Journal of Psychiatry*, *156*, 321–323.
Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.
Powell, R. A., & Gee, T. L. (1999). The effects of hypnosis on dissociative identity disorder: A reexamination of the evidence. *Canadian Journal of Psychiatry*, *44*, 914–916.
Prince, M. (1905). *The dissociation of a personality: A biographical study in abnormal psychology*. New York: Longmans, Green.
Putnam, F. W. (1989). *Diagnosis and treatment of multiple personality disorder*. New York: Guilford Press.
Putnam, F. W., Guroff, J. J., Silberman, E. K., Barban, L., & Post, R. M. (1986).

The clinical phenomenology of multiple personality disorder: Review of 100 recent cases. *Journal of Clinical Psychiatry*, *47*, 285–293.

Putnam, F. W., & Lowenstein, R. J. (2000). Dissociative identity disorder. In B. J. Sadock & V. A. Sadock (Eds.), *Kaplan and Sadock's comprehensive textbook of psychiatry* (7th ed., Vol 1, pp. 1552–1564). Philadelphia: Lippincott, Williams, & Wilkins.

Putnam, F. W., Zahn, T. P., & Post, R. M. (1990). Differential autonomic nervous system activity in multiple personality disorder. *Psychiatry Research*, *31*, 251–260.

Qin, J. J., Goodman, G. S., Bottoms, B. L., & Shaver, P. R. (1998). Repressed memory: An inquiry into allegations of ritual abuse. In S. J. Lynn (Ed.), *Truth in memory*. New York: Guilford Press.

Read, J. D., & Lindsay, D. S. (2000). "Amnesia" for summer camps and high school graduation: Memory work increases reports of prior periods of remembering less. *Journal of Traumatic Stress*, *13*, 129–147.

Reinders, A. A. T. S., Nijenhuis, E. R., Paans, A. M., Korf, J., Willemsen, A. T. M., & Den Boer, J. A. (2003). One brain, two selves. *NeuroImage*, *20*, 2119–2125.

Rifkin, A., Ghisalbert, D., Dimatou, S., Jin, C., & Sethi, M. (1998). Dissociative identity disorder in psychiatric inpatients. *American Journal of Psychiatry*, *155*, 844–845.

Rind, B., Tromovitch, P., & Bauserman, R. (1998). A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin*, *124*, 22–53.

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry*, *126*, 107–111.

Ross, C. A. (1990). Twelve cognitive errors about multiple personality disorder. *American Journal of Psychotherapy*, *44*, 348–356.

Ross, C. A. (1991). Epidemiology of multiple personality disorder and dissociation. *Psychiatric Clinics of North America*, *14*, 503–517.

Ross, C. A. (1994). *The Osiris complex: Case studies in multiple personality disorder*. Toronto, Canada: University of Toronto Press.

Ross, C. A. (1997). *Dissociative identity disorder: Diagnosis, clinical features, and treatment of multiple personality*. New York: Wiley.

Ross, C. A., & Ness, L. (2010). Symptom patterns in dissociative identity disorder patients and the general population. *Journal of Trauma and Dissociation*, *11*, 458–468.

Ross, C. A., & Norton, G. R. (1989). Effects of hypnosis on the features of multiple personality disorder. *Dissociation*, *3*, 99–106.

Ross, C. A., Norton, G. R., & Wozney, K. (1989). Multiple personality disorder: An analysis of 236 cases. *Canadian Journal of Psychiatry*, *34*, 413–418.

Sapolsky, R. M. (2000). Glucocorticoids and hippocampal atrophy in neuropsychiatric disroders. *Archives of General Psychiatry*, *57*, 925–935.

Sarbin, T. R. (1995). On the belief that one body may be host to two or more personalities. *International Journal of Clinical and Experimental Hypnosis*, *43*, 163–183.

Schreiber, F. R. (1973). *Sybil*. New York: Warner.

Schultz, R., Braun, B. G., & Kluft, R. P. (1989). Multiple personality disorder:

Phenomenology of selected variables in comparison to major depression. *Dissociation*, *2*, 45–51.
Scroppo, J. C., Drob, S. L., Weinberger, J. L., & Eagle, P. (1998). Identifying dissociative identity disorder: A self-report and projective study. *Journal of Abnormal Psychology*, *107*, 272–284.
Sekine, Y. U. Y. (2000). Dissociative identity disorder (DID) in Japan: A forensic case report and the recent increase in reports of DID. *International Journal of Psychiatry in Clinical Practice*, *4*, 155–160.
Shermer, M. (1997). *Why people believe weird things: Pseudoscience, superstition, and other confusions of our time*. New York: Freeman.
Shermer, M. (2001). *The borderlands of science: Where sense meets nonsense*. New York: Oxford University Press.
Showalter, E. (1997). *Hystories: Hysterical epidemics and modern culture*. New York: Columbia University Press.
Slovenko, F. (1999). The production of multiple personalities. *Journal of Psychiatry and Law*, *27*, 215–253.
Sno, H. N., & Schalken, H. F. (1999). Dissociative identity disorder: Diagnosis and treatment in the Netherlands. *European Psychiatry*, *5*, 270–277.
Southwick, S., Morgan, A. C., Nicolaou, A. L., & Charney, D. S. (1997). Consistency of memory for combat-related traumatic events in veterans of Operation Desert Storm. *American Journal of Psychiatry*, *154*, 173–177.
Spanos, N. P. (1994). Multiple identity enactments and multiple personality disorder: A sociocognitive perspective. *Psychological Bulletin*, *116*, 143–165.
Spanos, N. P. (1996). *Multiple identities and false memories: A sociocognitive perspective*. Washington, DC: American Psychological Association.
Spanos, N. P. & Chaves, J. F. (1989). *Hypnosis: The cognitive-behavioral perspective*. Buffalo, NY: Prometheus.
Spanos, N. P., Weekes, J. R., & Bertrand, L. D. (1985). Multiple personality: A social psychological perspective. *Journal of Abnormal Psychology*, *94*, 362–376.
Spanos, N. P., Weekes, J. R., Menary, E., & Bertrand, L. D. (1986). Hypnotic interview and age regression procedures in the elicitation of multiple personality symptoms. *Psychiatry*, *49*, 298–311.
Spiegel, D. (1993, May 20). Letter to the Executive Council, International Study for the Study of Multiple Personality and Dissociation. *News, International Society of the Study of Multiple Personality and Dissociation*, *11*, 15.
Stafford, J., & Lynn, S. J. (1998). *Cultural scripts, childhood abuse, and multiple identities: A study of role-played enactments*. Manuscript submitted for publication.
Stein, M. B., Koverola, C., Hanna, C., & Torchia, M. G. (1997). Hippocampal volume in women victimized by childhood sexual abuse. *Psychological Medicine*, *27*, 951–959.
Takahashi, Y. (1990). Is multiple personality disorder really rare in Japan?: Dissociation. *Progress in the Dissociative Disorder*, *3*, 57–59.
Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, *83*, 268–277.
Thigpen, C. H., & Cleckley, H. M. (1957). *The three faces of Eve*. New York: McGraw-Hill.

Trifonova, T. (2010). Multiple personality and the discourse of the multiple in Hollywood cinema. *European Journal of American Culture*, *29*(2), 145–171.

Tsai, G. E., Condie, D., Wu, M.-T., & Chang, I.-W. (1999). Functional magnetic resonance imaging of personality switches in a woman with dissociative identity disorder. *Harvard Review of Psychiatry*, *72*, 119–122.

van der Hart, O. (1993). Multiple personality disorder in Europe: Impressions. *Dissociation*, *6*, 102–118.

van der Kloet, D., Giesbrecht, T., Lynn, S. J., Merckelbach, H., & de Zutter, A. (2012). Sleep normalization and decrease in dissociative experiences: Evaluation in an inpatient sample. *Journal of Abnormal Psychology*, *121*, 140–150.

van der Kolk, B. A., van der Hart, O., & Marmar, C. R. (1996). Dissociation and information processing in posttraumatic stress disorder. In B. A. van der Kolk, A. C. McFarlane, & L. Weisaeth (Eds.), *Traumatic stress: The effects of overwhelming stress on mind, body, and society*. New York: Guilford Press.

Vermetten, E., Schmahl, C., Lindner, S., Loewenstein, R. J., & Bremner, J. D. (2006). Hippocampal and amygdalar volumes in dissociative identity disorder. *American Journal of Psychiatry*, *163*, 630–636.

Walker, H. (2008). *Breaking free*: *My life with dissociative identity disorder*. New York: Touchstone.

Waller, N. G., & Ross, C. A. (1997). The prevalence and biometric structure of pathological dissociation in the general population: Taxometric and behavior genetic findings. *Journal of Abnormal Psychology*, *106*, 499–510.

Widom, C. S. (1989). The cycle of violence. *Science*, *244*, 160–166.

Wilson, N. L. (2003). Commercializing mental health issues: Entertainment, advertising, and psychological advice. In S. O. Lilienfeld, J. M. Lohr, & S. J. Lynn (Eds.), *Science and pseudoscience in contemporary clinical psychology* (pp. 425–459). New York: Guilford Press.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., & Dong, Q. (in press). The relationship between DRM and misinformation false memories. *Memory and Cognition*.

Zohar, J. (1998). Post-traumatic stress disorder: The hidden epidemic of modern times. *CNS Spectrums*, *3*(7, Suppl. 2), 4–51.

PART II

OVERARCHING CONTROVERSIES IN PSYCHOLOGICAL TREATMENT

CHAPTER SIX

The Science of Psychotherapy

Developing, Testing, and Promoting Evidence-Based Treatments

Brandon A. Gaudiano, Kristy L. Dalrymple,
Lauren M. Weinstock, and Jeffrey M. Lohr

"Psychotherapy" is a generic term that encompasses a wide variety of approaches with disparate philosophies, theoretical frameworks, putative mechanisms of change, techniques, and strategies. Hundreds of different psychotherapies have been developed over the years (Singer & Lalich, 1996; Wampold, 2001), and the number is continuing to rise. However, not all psychotherapies are created equal. Some represent vestiges from psychotherapy's prescientific past (e.g., Freudian psychoanalysis; Cioffi, 1998), others represent approaches that have been shown to be as or more effective than powerful psychotropic medications for many common emotional disorders (e.g., cognitive-behavioral therapies; Butler, Chapman, Forman, & Beck, 2006), and still others may cause more harm than good (e.g., psychological "debriefing"; Lilienfeld, 2007). Finally, other forms of psychotherapy strain credulity and propose techniques for treating psychological disorders that have little connectivity with the science of psychology or biology (e.g., "energy meridian therapies"; Gaudiano & Herbert, 2000).

What explains this motley crew of psychotherapies? One answer can be found in the tendency of many therapists to choose the treatments that they use based on clinical intuition and personal experience rather than scientific support (Gaudiano, Brown, & Miller, 2011; Stewart & Chambless, 2007). For example, a survey of 891 psychologists showed that 47% reported never using evidence-based psychotherapy treatment manuals in

their clinical practice, and only 6% reported using them often (Addis & Krasnow, 2000). Another survey of practicing psychologists indicated that, although most reported using cognitive-behavioral strategies in general when treating anxiety, the majority of clinicians were not using exposure techniques that many consider the essential ingredient of effective treatment (Freiheit, Vye, Swan, & Cady, 2004). Conversely, many clinicians appear to be using unsupported therapy techniques. For example, Gaudiano, Brown, and Miller (2012) surveyed 149 community therapists and found that 42% reported that they use or are inclined to use unsupported "energy meridian therapy" techniques. Another survey of 191 social workers found that 76% reported using at least one "novel unsupported therapy" (e.g., reparenting therapy, age regression therapy for sexual abuse) during the past year (Pignotti & Thyer, 2009a).

However, there is another path to understanding and practicing psychotherapy. In 1967, psychologist Gordon Paul issued his challenge to the field when he asked, "What treatment, by whom, is most effective for this individual with that specific problem, and under which set of circumstances?" (p. 111). Psychotherapy researchers have been seeking answers to this important question ever since. The results of this research have been a boon to society and have yielded many highly effective treatments for conditions as diverse as autistic disorder, major depression, social phobia (especially severe public-speaking anxiety), insomnia, chronic pain, posttraumatic stress disorder, and schizophrenia (when combined with medication, as psychotherapy alone is not recommended for schizophrenia). In contrast, the societal costs of unsupported and pseudoscientific psychotherapies in terms of ineffective and even harmful treatments are vast. However, psychotherapy research serves as an antidote to pseudoscientific practices by providing the tools necessary to separate empirically supported from unsupported practices to identify the most effective treatments for consumers of mental health services. In this chapter, we provide the reader with a general overview of the process of psychotherapy research, including the development of new therapies, testing under controlled conditions, evaluations in real-world settings, and dissemination efforts so that evidence-based therapies will be made more widely available to the general public.

HISTORY OF PSYCHOTHERAPY RESEARCH

Before describing the intricacies of psychotherapy research, we will highlight the historical evolution of psychotherapy research by describing some of the seminal studies that shaped the research process. Although a plethora of scientific articles is now available on the testing of psychotherapies, there was a time when such articles were few and far between. However, over the past 60 years, psychotherapy has been subjected to more rigorous examinations, and the methodology of this scientific endeavor has been

refined to advance our understanding of how best to help individuals suffering from a range of problems.

A review published by Eysenck in 1952 spurred this movement of more systematically examining the efficacy of our treatments. In this article, the first critical review of psychotherapy, Eysenck (1952) reviewed outcome data from over 7,000 patients who had received psychoanalytic or eclectic therapy. For comparison purposes, he reviewed hospital discharge rates and improvement of patients receiving medical treatment from general practitioners. Based on his review, he found that only 44% of those receiving psychoanalytic therapy and 66% of those receiving eclectic therapy improved, compared with 72% of people in the comparison group. Thus, he concluded that there was no scientific evidence for the efficacy of psychotherapy. Not surprisingly, Eysenck's review was not well received by many in the field, and it sparked much controversy and criticism. Among the many criticisms, the more central ones focused on his lack of a true control group, his failure to match participants across treatment and comparison conditions, and the methods he used for calculating improvement rates (e.g., Bergin, 1971). As a result, researchers began conducting formal research to test psychotherapies in an effort to provide support for the efficacy of psychotherapy.

Eysenck went on to publish other critical reviews in 1961 and 1966 that reiterated his conclusion that there was a lack of evidence for the efficacy of psychotherapy. However, he also refined this conclusion by arguing that therapies based on learning theory, namely, behavior therapies, were efficacious. These latter reviews raised questions about the comparative efficacy of different types of psychotherapies. Prior to Eysenck's earlier review, many of the psychotherapies being practiced were insight-oriented and based on Freudian principles. However, behavior modification and other types of psychotherapies were being developed and popularized in the 1950s and 1960s, making it more important to begin examining whether some types of psychotherapies were more efficacious than others. Therefore, the focus shifted from a question of whether psychotherapy was efficacious in general to a question of which therapies were efficacious relative to others.

One of the earlier comparative studies was conducted at Temple University (Sloane, Staples, Cristol, Yorkston, & Whipple, 1975). In this study, 90 patients were assigned to behavior therapy, psychoanalytic therapy, or wait-list control (i.e., when subjects are assessed for a period of time prior to the initiation of treatment to obtain a no treatment baseline). The methods included several procedures commonly used to this day, such as use of pre- and posttreatment assessments, standardized interviews, informant reports, and data from multiple sources (e.g., self-report, clinician-rated, and independent assessor-rated). Results showed that although there was a slight trend in favor of behavior therapy overall, for the most part there was no difference between the two active treatments. Patients in the wait-list control group also improved, but less so than patients undergoing the two active treatments.

Another seminal comparison study was the National Institute of Mental Health's Treatment of Depression Collaborative Research Program (NIMH TDCRP; Elkin et al., 1989), which is considered one of the most extensive and methodologically sound comparative trials. The study compared cognitive-behavioral therapy (CBT), interpersonal psychotherapy (IPT), an antidepressant medication (imipramine) and a placebo control for depression. Consistent with the Temple study, results from this study showed no initial difference in outcomes between the three active treatments at posttreatment. However, there has been some controversy regarding these results. Stating that the original report used too stringent criteria for statistical significance, Klein and Ross (1993) reanalyzed that data. They found that particularly among patients with severe depression, medication was superior to psychotherapy and IPT was superior to CBT. Others have disagreed with this reanalysis (e.g., Jacobson & Hollon, 1996), noting site differences (and thus possible allegiance effects) as potential sources of bias in the results. Blatt, Sanislow, Zuroff, and Pilkonis (2000) reanalyzed the data once again, finding no differences between active treatments at posttreatment and 18-month follow-up based on symptom reduction. However, they found differences in life adjustment: Patients who received IPT reported greater satisfaction with treatment, and patients who received IPT and CBT reported greater effects on their capacity to (1) establish and maintain interpersonal relationships and (2) recognize and understand the sources of their depression relative to those in the medication and placebo conditions.

The "Dodo Bird" Verdict

These and other earlier studies that failed to find differences between types of psychotherapies renewed interest in what is referred to as the "dodo bird" verdict (Luborsky, Singer, & Luborsky, 1975). This concept is based on a scene in *Alice in Wonderland*, in which after a race the dodo bird states, "Everyone has won, and all must have prizes." This intepretation was first applied to psychotherapy in 1936 by Rosenzweig, who proposed that all psychotherapies share common factors that are used by therapists to foster change in clients, and thus all psychotherapies are virtually the same. However, this hypothesis remained as such because at that time there was no psychotherapy research to provide supporting evidence and psychotherapies consisted of variants of one model, namely, the psychoanalytic model.

The development of other psychotherapy models (e.g., behavior therapy) in subsequent decades prompted comparative outcome studies such as the NIMH TDCRP (Elkin et al., 1989), the results of which refueled the dodo bird verdict debate. Some meta-analyses also have suggested that there are few differences between psychotherapies (e.g., Smith & Glass, 1977; Wampold et al., 1997), although these studies have been criticized

for their methodology. For example, the Smith and Glass (1977) meta-analysis (see the section "Meta-Analysis" below) categorized cognitive therapies as "nonbehavioral" interventions (along with psychodynamic and humanistic therapies), included few well-controlled studies of behavior therapy, and did not control for the quality of studies that were included in the meta-analysis (Rachman & Wilson, 1980). Since these earlier meta-analyses, other reviews have attempted to account for these limitations. For instance, Hunsley and Di Giulio (2002) reanalyzed data from Smith and Glass (1977) to correct for methodological concerns and found that the behavioral therapies had superior outcomes relative to other types of therapies. Tolin (2010) conducted a meta-analysis comparing CBT to other psychotherapies, accounting for other factors such as study quality, outcome measures used, age and diagnosis of the samples, and researcher allegiance. Researcher allegiance was assessed by contacting the principal investigators of the studies and having them rate their allegiance to CBT versus the alternative treatment comparison on a scale ranging from +3 (strongly favored [name of CBT therapy]) to –3 (strongly favored [name of alternative therapy]). Allegiance was defined for the investigators as "the extent to which you believe in a treatment, expect it to succeed beyond the nonspecific effects of treatment, or identify yourself as a proponent of that treatment." Across 26 studies and a sample of 1,981 patients, Tolin found that CBT was superior to psychodynamic therapy at posttreatment and at follow-up, particularly for patients with anxiety and depressive disorders. CBT remained superior even after controlling for researcher allegiance effects. However, CBT was not superior to interpersonal or supportive psychotherapies. Therefore, these more sophisticated studies have called into question the notion that all therapies are created equal. However, a follow-up meta-analysis contradicted this conclusion (Baardseth et al., 2013).

Thus, the dodo bird debate continues, with proponents suggesting that all psychotherapies are equally efficacious due to common factors (e.g., therapeutic alliance), and critics arguing that some specific interventions are more helpful than others for particular disorders and that psychotherapy "common factors" may be necessary but not sufficient facilitators of change (see Budd & Hughes, 2009, for a review). Despite this ongoing debate, it has been acknowledged more recently that perhaps a more fruitful endeavor would be to examine the processes through which these treatments produce change, rather than to determine how each of these treatment packages compares to one another in overall efficacy (Budd & Hughes, 2009; Kazdin, 2007; Mansell, 2011).

Thus, the research question has shifted once again to ask whether, within a given treatment, certain components may be superior to others. This focus has prompted a methodology termed "component analysis" or "dismantling" studies, in which specific components of a treatment are isolated and tested against the "full-package" treatment. For example, some CBT researchers began to question whether the cognitive component of

treatment was superior to the behavioral component, and vice versa. In a seminal study, Jacobson and colleagues (1996) compared the behavioral activation component (consisting of monitoring and assigning daily activities in order to engender a sense of mastery or pleasure) with behavioral activation plus modification of automatic thoughts (but not modification of core schemas) to full-package CBT. The study included 150 participants diagnosed with major depression who were randomly assigned to one of those three conditions. Results showed significant improvements over the course of treatment, with no difference between the components and full-package. Interestingly, they also found that behavioral activation alone altered negative thinking just as effectively as the other treatments that more directly attempted to address cognitions. In recent decades, this study has sparked greater interest in testing the proposed mechanisms of action of various psychotherapies.

Psychotherapy research began as a response to a scathing review regarding the lack of evidence for its benefit (Eysenck, 1952). Over time, the goal of psychotherapy research evolved; it no longer was enough to provide support for whether or not psychotherapy in general was beneficial. With the development of various types of psychotherapies, it became important to know whether some treatments were more efficacious than others. After some studies suggested that there may not be large differences among types of psychotherapies in general, the research question once again evolved: Are there particular aspects of specific therapies that contribute more to change than others? Therefore, over the decades the field of psychotherapy research has attempted to come closer to answering Gordon Paul's (1967) "ultimate" clinical question.

MODELS OF PSYCHOTHERAPY TREATMENT DEVELOPMENT

A psychotherapy must first be developed before it can be tested under scientific conditions. Traditionally, psychotherapies have been (and many still are) developed based solely on clinical insights gleaned from the therapist's personal experience with treating patients and his/her exposure to various theories and practices in the field. However, it is important to differentiate between what Reichenbach (1938) called the "context of discovery" versus the "context of justification." Although clinical insights may constitute an important element in the initial *development* of a new psychotherapy, they are insufficient for determining whether the treatment is *efficacious* or how it should be modified to ensure safety and improve outcomes. There is no one universal model of treatment development favored by all psychotherapy researchers. However, Rounsaville, Carroll, and Onken (2001) proposed a "stage model" of development that is currently favored by the National Institutes of Health for developing and testing novel psychosocial treatments. In their model, Stage I consists of various development activities, including writing the treatment manual, producing the therapist training

program, creating the therapist fidelity and competence scales for ensuring reliable and proficient administration of the treatment, and pilot testing the intervention to establish feasibility, patient acceptability, and initial efficacy. Stage II entails formal efficacy research to establish the effects of the intervention under tightly controlled experimental conditions using large samples, and Stage III focuses on effectiveness testing under real-world conditions and broader dissemination efforts (see below for more information on efficacy versus effectiveness research).

In the Rounsaville et al. (2001) model, each stage is broken down into various substages and activities. In Stage I, the development of a new psychotherapy should be based on a sound psychological theory and incorporate findings from previously conducted basic behavioral research to inform the treatment processes and strategies. Furthermore, the process is iterative and incremental in that the treatment is initially developed and pilot tested in an open trial (i.e., an uncontrolled pre–post study design) to obtain experience conducting the treatment and to consider modifications and revisions based on feedback from patients and therapists. Once the treatment protocol is further refined and elaborated, a pilot randomized controlled trial is conducted in which a smaller sample of patients is randomly assigned to different conditions and the control condition is also pilot tested to ensure its feasibility and acceptability. Also prior to conducting the pilot clinical trial, the researchers develop therapist fidelity/competence scales to measure how well study therapists adhere to the treatment manual and show skill in implementing the procedures. These scales are then tested during the pilot clinical trials. Following the completion of the open and randomized pilot studies, the treatment manual, training program, and fidelity/competence scales are further refined in preparation for a full-scale randomized control trial as part of Stage II efficacy testing.

The above-mentioned model is logical and relatively straightforward, but it has notable limitations. Although the process is proposed to be flexible and iterative, treatments based on the stage model tend to be focused on specific DSM diagnoses and are not necessarily representative of the broader array of patients treated by clinicians in routine practice. In addition, treatment manuals often become too narrowly focused on a specific clinical problem, and therapists treating a multi-problem patient are left to decipher on their own how to incorporate several different manuals in treatment. Westen and colleagues (2004), though acknowledging that treatment manuals are necessary for conducting methodologically sound psychotherapy research, argued that the way in which these treatment manuals are implemented limits their utility and flexibility in clinical practice. In addition to inadequately addressing comorbidity, they noted that treatment manuals tend to minimize the use of clinical judgment and thus have transformed from being descriptive (i.e., operationalizing the treatment) to prescriptive (i.e., outlining specific procedures that should be followed). Not surprisingly, some treatments developed from the stage model have run into problems at later stages of testing when attempts are made to implement

them in real-world treatment settings, and fundamental issues of feasibility are not sufficiently addressed. Furthermore, the stage model has tended to create complex, multi-package treatments with unclear mechanisms of action because the active versus superfluous ingredients of the treatment are not investigated until the later stages. In addition, Kraemer, Mintz, Noda, Tinklenberg, and Yesavage (2006) urged caution when interpreting effect sizes derived from underpowered Stage I pilot studies because they often are unreliable due to error variance and do not accurately predict effects obtained in future larger-scale randomized trials of the treatment. Cohen (1988) described conventions for interpreting effect sizes for group comparisons (*d*s), with 0.20 representing a small effect, 0.50 a medium effect, and 0.80 a large effect

These limitations have led some authors to propose alternative models for developing and testing novel psychosocial interventions. For example, Hayes, Long, Levin, and Follette (2013) recently proposed what they call a contextual behavioral science approach to treatment development. They argue for a more "reticulated" or networked approach with greater: (1) philosophical clarity and specification of underlying assumptions, (2) distinction between clinical versus basic models and their relationship, (3) targeting of manipulable contextual variables, (4) attention to mechanisms of action issues early on, and (5) focus on dissemination and public health impact at the beginning and throughout the process. They also suggest that treatment development requires a community of applied and basic researchers conducting a series of different types of studies over time and working together by finding mutual interests.

EFFICACY RESEARCH

Once a treatment is developed, it is tested under controlled conditions to draw cause-and-effect conclusions. When evaluating the question of whether psychotherapy—or any particular psychotherapy—"works," it is important to acknowledge that different clinical research methodologies can be used to answer different, albeit complementary, research questions. One such methodology is the efficacy trial, which seeks to answer questions about whether a treatment will work under the most controlled conditions. Efficacy research represents a fundamental step in the evaluation of a treatment in that it is used to demonstrate whether a treatment *can* produce a desired effect. If a treatment is shown to effect change on a targeted outcome under highly controlled conditions, the next step would be to examine the same treatment under more clinically heterogeneous circumstances (i.e., in an effectiveness trial). As such, efficacy research maximizes *internal validity*, which is the extent to which one can rule out alternative explanations for treatment outcomes, over *external validity*, which is the extent to which the treatment outcomes can be generalized to "real-world" clinical settings and populations (Nathan, Stuart, & Dolan, 2000).

The randomized controlled trial (RCT) represents the "gold standard" study design in efficacy research. In an RCT, study participants are randomly assigned to either the treatment of interest or a comparison condition, and then assessed at regular intervals to determine differential outcome by group. Within this general framework, there are several key research design elements that are considered ideal in an RCT to evaluate *psychotherapy efficacy* (cf. Kraemer, Wilson, Fairburn, & Agras, 2002). First, an efficacy trial benefits from carefully selected, diagnostically homogeneous samples, so as to eliminate potential "third variables" of a clinical nature that might unduly influence treatment outcome. Second, the sample should be of adequate size to have statistical power to detect clinically significant differences between study conditions and to provide accurate effect size estimates (i.e., a measure of the magnitude of the improvement observed from the treatment). Third, an efficacy trial makes use of highly operationalized treatment and comparison control conditions, which are typically specified in a manual or other protocol document. This approach minimizes heterogeneity in the delivery of the treatment of interest and in its comparator, ensures that participants in the control condition are not inadvertently exposed to the "active ingredients" of the experimental treatment, and allows for replication in both research and clinical settings. Fourth, efficacy trials rely on highly trained, highly supervised therapists, who are regularly assessed for adherence to the study treatment and for competence in their delivery. As such, when evaluating study outcomes, one can be confident that the treatment of interest was delivered properly and was of high quality. Fifth, it is critical that study investigators and clinical raters be "blind" to—or unaware of—participant assignment to study condition, so as not to bias outcomes through expectancies or allegiances to the treatment(s) being studied. Ideally, it is important that study participants be blind to treatment assignment for these same reasons, yet this requirement is rarely practical in psychotherapy research due to the nature of the treatments studied and the comparison conditions employed.

Indeed, the choice of a comparator in an efficacy trial will be largely driven by the clinical question of interest (Kazdin, 1998; Nathan et al., 2000). If one is interested simply in whether a treatment produces therapeutic change, a no-treatment, treatment as usual, or wait-list control condition might be used. However, in psychotherapy research, one concern may be that those receiving the treatment of interest may improve as a result of increased clinical attention; in which case an "active control" condition that proscribes elements of the treatment of interest may be used as an alternative comparison condition. This active control condition may be a sham intervention designed specifically for the research study to account for some "nonspecific" effects (Imel & Wampold, 2008; see the section "Placebo and Nonspecific Factors" below) of psychotherapy treatment or an existing treatment with established efficacy itself. This latter study design is often referred to as a *comparative efficacy trial*. Still other RCT designs may be used to evaluate the efficacy of a treatment when it is delivered in

addition to an existing treatment with established efficacy. In these adjunctive treatment designs, a single treatment will be compared against some combination of treatments, such as a psychotherapy delivered alone versus the same psychotherapy delivered as an adjunct to medication. Finally, others might be interested in using an RCT to "dismantle" a treatment and evaluate the relative efficacy of its different components. One of the more notable examples of this efficacy design is the study conducted by Jacobson and colleagues (1996) described earlier.

Mediators and Moderators of Treatment Effects

Although the efficacy research design allows one to examine the question of whether a treatment works, it does not necessarily answer the question of *how* it works. There has been increasing interest in the mechanisms through which treatments enact their response, often referred to as *mediators* of treatment (Kazdin, 2007). In their seminal paper, Baron and Kenny (1986) described a set of criteria for the establishment of mediation, in that a treatment (*T*) results in a particular outcome (*O*) through change in the mediating variable (*M*). More recently, Kraemer et al. (2002) and Kraemer, Kiernan, Essex, and Kupfer (2008) clarified the theoretical and the statistical requirements for establishing mediation, the conventions for which have been referred to as the MacArthur Approach. Identifying the mediators of treatment can be important for several reasons. Knowledge of mediators can prompt researchers to strengthen, add, or remove certain intervention components to improve clinical outcomes following multicomponent psychotherapy protocols (Kraemer et al., 2002; Kazdin, 2007). A better understanding of the mechanisms of treatment also can facilitate their use in clinical practice by providing knowledge on the components that are most likely to effect change (Kazdin, 2007). Identification of mediators can also lead to newer treatments. For example, in the CBT dismantling study conducted by Jacobson et al. (1996), there was a paradoxical finding in which the effect of the cognitive interventions on depressive symptoms was mediated through change in behavioral activation (BA), whereas the effect of the behavioral intervention on depressive symptoms appeared to be mediated through change in cognition. Jacobson and colleagues concluded that BA was a more parsimonious treatment than the full CBT package that yielded similar clinical outcomes via the same mechanism (i.e., cognitive change) as CBT. Following from these findings, Jacobson and others went on to develop updated, stand-alone BA interventions for depression that have since become the target of a large body of efficacy research (e.g., Dimidjian et al., 2006; Hopko, Lejuez, Ruggiero, & Eifert, 2003).

In addition to elucidating how psychotherapy treatments work, researchers and clinicians have become increasingly interested in *who* is more or less likely to respond to treatment or *under what conditions* a treatment is likely to be most efficacious. Such characteristics are often referred to as *moderators* of treatment (e.g., sex, race/ethnicity, diagnosis,

family functioning). Similar to mediation, criteria for the identification and evaluation of treatment moderators were operationalized by Baron and Kenny (1986) and expanded on by Kraemer et al. (2002, 2008) in their MacArthur Approach. The identification of moderators is important for several reasons. First, moderators can provide information to clinicians as to whether a treatment is likely to be useful for a particular patient and whether an alternative treatment might be indicated (Kazdin, 2007). Second, researchers can use moderators to optimally target treatment to certain populations who are more likely to respond, thereby increasing the power to evaluate treatment efficacy in subsequent RCTs. Despite the clear clinical applicability and empirical interest in identification of psychotherapy moderators, research attempting to do so has been somewhat mixed (Roth & Fonagy, 1996). For example, no benefit was found in matching patients with alcoholism to treatment in the large and costly Project MATCH (Matching Alcoholism Treatment to Client Heterogeneity, Project MATCH Research Group, 1997).

EFFECTIVENESS RESEARCH

In contrast to efficacy studies, which examine therapies under tightly controlled conditions, effectiveness studies attempt to determine to what degree treatment can improve outcomes in routine practice under more naturalistic conditions. Thus, generally the focus of effectiveness studies is on external validity. Although efficacy studies are critical to psychotherapy research, critics have argued that efficacy studies alone are not sufficient in determining what treatments can improve outcomes. One such criticism has been that results from efficacy studies may not translate to patients who present to routine practice for treatment, given that patients in these "real-world" settings have greater comorbidity and severity compared with patients typically enrolled in efficacy trials (Westen & Morrison, 2001). Treatment preference also may play a role in that patients in efficacy trials are randomly assigned to a treatment condition; if they receive a treatment that they do not prefer, this could have an effect on outcome that would not be seen in routine settings, where patients have a choice of treatment (Iacoviello et al., 2007). However, other research shows that treatment preference is not always related to outcome (e.g., Leykin et al., 2007).

Other criticisms of efficacy studies focus on practical factors of conducting research under usual care conditions. For example, a hallmark of efficacy trials is the use of a specific treatment manual, with special training for the therapists and ongoing monitoring. Due to resource (e.g., monetary) and time (i.e., higher caseloads) constraints in routine practice settings, it may be difficult, if not impractical, to have the same level of methodological rigor as would be the case in efficacy trials (Schoenwald & Hoagwood, 2001).

Effectiveness studies thus arose out of these concerns, in addition to

concerns of small sample size and low statistical power in efficacy studies. Therefore, effectiveness studies generally consist of larger samples, use quasi-experimental designs, and evaluate long-term and broader outcomes (Wells, 1999). These features also mean that certain drawbacks are present with effectiveness designs: There tends to be greater heterogeneity in implementing the treatment; methodology tends to be less rigorous compared with efficacy trials, resulting in lower internal validity; and although longer-term follow-up assessments are conducted, they tend to be less comprehensive compared with efficacy studies (Wells, 1999).

Despite these limitations, effectiveness studies have provided important information as to how well efficacious treatments translate to "real-world" practice. Seligman (1995) credited a survey by *Consumer Reports* in 1994 as having pioneered the effectiveness methodology. Out of 180,000 readers who received this annual survey, 7,000 responded to the mental health questions, with 2,900 reporting that they had seen a mental health professional in the past 3 years (including psychologists, psychiatrists, social workers, or marriage counselors). Respondents answered questions related to their reason for seeking treatment, what type of professional they saw, the type of treatment they received, and reasons for termination. Overall, results showed that: (1) most (approximately 90%) of the respondents reported having experienced improvement; (2) long-term therapy appeared to produce greater improvement compared with short-term therapy; (3) psychotherapy alone did not differ from psychotherapy plus medication; and (4) no specific psychotherapy outperformed the others (Seligman, 1995).

The *Consumer Reports* study possessed several strengths, most importantly that it reflected treatment as it is administered in the community. Nevertheless, this study was also marked by limitations, including retrospective reporting of baseline and posttreatment emotional states, reliance on unverified self-reports, and lack of a control group. Therefore, it cannot be determined from this study whether people who reported benefit from treatment would have improved simply as a result of the passage of time or of some other intervention (e.g., talking to friends). The various limitations posed by both efficacy and effectiveness studies highlight the need for both types of methodologies in psychotherapy research. Efficacy studies can provide the methodological rigor necessary to determine whether a type of therapy outperforms other treatments, whereas effectiveness studies can determine to what degree those treatments are feasible and beneficial in typical clinical practice.

HYBRID EFFICACY–EFFECTIVENESS RESEARCH

More recently, emphasis has been placed on designing "hybrid" studies that incorporate efficacy and effectiveness methodology. Such strategies may include use of a large, practice-based sample for assessing effectiveness,

as well as use of a subsample of this population to conduct an efficacy study. In addition, these strategies may include initially training therapists with a manual but not using that manual during the course of the study, and comparing a specific treatment with treatment as usual in that clinical practice (Clarke, 1995). Furthermore, Wells (1999) suggested that hybrid study designs could include testing efficacious treatments or approximations of those treatments using typical providers and patient populations, collecting data on outcomes as well as societal and cost-effectiveness data, and identifying predictors of treatment adherence rather than simply dropping nonadherent patients from analyses.

In 1994, the Pennsylvania Psychological Association formed a task force to develop a Psychotherapy Research Network (PRN), with the aim of producing an infrastructure to conduct effectiveness research in a scientifically valid manner (thus creating a hybrid study of efficacy and effectiveness). Borkovec, Echemendia, Ragusea, and Ruiz (2001) described the development of the PRN and presented data on some of the initial results. The PRN was developed by a committee of both practitioners and scientists, in an attempt to address concerns associated with both research and clinical practice. Practitioners volunteered to take part in the PRN and obtained informed consent from their patients. Their patients completed pre-, mid-, and posttreatment assessments. Preliminary data were analyzed from an initial feasibility trial, in which 77 therapists participated. The therapists had an average caseload of 23 clients per week, and most (approximately three quarters) had a Ph.D. Initial assessments were obtained from 220 clients; 75 completed the midtreatment assessment, and 42 completed the posttreatment assessment. Results showed significant decreases in symptoms and significant improvement in functioning from pre- to midtreatment and pre- to posttreatment, although functioning was more globally improved at posttreatment compared with midtreatment. The initial findings from this trial were encouraging and indicated that more rigorous research can be applied in routine practice settings.

Since the development of the PRN, other researchers have utilized a hybrid methodology to examine the effectiveness of CBT in practice settings. For instance, Westbrook and Kirk (2005) used a "benchmarking" method, in which they examined outcome data from patients who received uncontrolled treatment (e.g., CBT) within a routine clinical setting, and the results were compared with similar data from efficacy studies. From a sample of 1,276 patients who completed treatment, effect size changes in depression and anxiety were moderate in magnitude when including the total sample and large when excluding patients whose baseline scores were already within the normal range. Half of the sample showed clinically significant and reliable improvement, and 30% of patients met the criteria for recovery. When benchmarking results from their trial to other previous trials, Westbrook and Kirk (2005) found that the effect sizes for anxiety were somewhat lower compared with efficacy trials, but depression

effect sizes were comparable. McEvoy and Nathan (2007) also conducted a benchmarking study for mixed-diagnosis (depression and anxiety) patients receiving CBT in a community mental health clinic in Australia. Unlike the Westbrook and Kirk (2005) study, this study utilized a diagnostic interview and manual-based treatment. Out of 143 patients, the effect size was large for depression and moderate for anxiety. Although their effect sizes were somewhat lower compared with efficacy trials, results were within the range of effect sizes from those trials. Given the large amount of evidence for the efficacy of psychotherapy, particularly CBT, there has been a call to move into a new phase of psychotherapy research in which the "transportability" of these efficacious treatments is examined in typical clinical settings. Effectiveness studies provide important information as to how efficacious treatment may apply to the typical patient in routine clinical practice.

META-ANALYSIS

Along with the proliferation of psychotherapy treatment research, there has been significant interest in compiling data across studies in an effort to synthesize, and at times reconcile, findings concerning the efficacy and effectiveness of our treatments. Historically, such synthesis has been conducted by means of a narrative review, in which a scholar relies on a largely qualitative process to compile, review, and analyze a body of research that addresses a specific question, and then formulates conclusions concerning the nature of the evidence. Yet this narrative review process has been criticized as too subjective and vulnerable to bias, both in the selection of studies to be included in the review and in the resulting conclusions drawn (Glass, 1976). It has been further argued that the narrative review is an inexact process through which one can evaluate the magnitude and probability of treatment effects (Glass, 1976), given its qualitative nature.

In response to the limitations of the narrative review, Smith and Glass (1977) and their colleagues (Glass et al., 1981; Smith, Glass, & Miller, 1980) developed a set of quantitative procedures by which effect size estimates, or statistical calculations representing the magnitude of treatment effects, are derived from multiple studies and then translated into a common effect size for interpretation. This quantifiable metric, along with its variability across studies, is then used to generate conclusions regarding the magnitude, pattern, and direction of treatment outcomes. As an "analysis of analyses" (Glass, 1976, p. 3), the process of *meta-analysis* combines many different effect sizes across many different studies, thereby providing better estimates of population outcomes than any one investigation. Further, by compiling data from multiple settings, samples, assessments, and methodologies, questions can be asked of the literature that any one study cannot easily provide. As such, not only does meta-analysis allow one to formulate conclusions regarding a treatment's effect, but it can be used to evaluate new research questions and elucidate novel directions for

future research (Bangert-Drowns, 1995). In their seminal meta-analysis of the psychotherapy outcome literature, which compiled data from over 400 research studies, Smith and Glass (1977) and Smith et al. (1980) evaluated the overall efficacy of psychotherapy treatment, as well as factors that might impact differential efficacy (e.g., type of therapy, type of outcome, length of treatment, therapist experience). With an average effect size of 0.85 (Smith et al., 1980), revealing that 80% of those who received a psychotherapy intervention were more improved than those who remained untreated, they concluded that psychotherapy produces meaningful clinical change. Notably, this study included all published trials for which effect sizes could be calculated, regardless of treatment type. When the meta-analysis was repeated after excluding those treatments that were considered to be undifferentiated, the effect size estimate increased to 0.93.

Smith et al.'s (1980) work yielded a number of additional influential findings. Magnitude of treatment outcome did not differ as a function of theoretical orientation (e.g., psychodynamic, cognitive, Rogerian), the psychotherapy process (e.g., verbal, behavioral, expressive), length of treatment, or treatment modality (e.g., individual vs. group). Similarly, therapist experience did not relate significantly to the magnitude of treatment effects. There was some evidence that effect sizes were largest for treatments for depression, simple (now termed specific) phobias, and their analogues, regardless of orientation or modality. Yet the generally strong effects of psychotherapy appeared to decrease over time within clients, with an average effect size of 0.50 at 2 years posttreatment.

Given the provocative nature of these data and the fact that this study represented the first application of meta-analysis in the clinical research literature, a number of critiques emerged that not only addressed the limitations of this meta-analysis, but also highlighted the limitations of meta-analysis more generally. The primary criticism leveled against this work was that it was too inclusive, relying on data from studies varying in scientific rigor. It was argued that any methodological flaws in the primary studies could differentially influence the meta-analytic outcomes themselves. Although Smith et al. (1980) accounted for the quality of the study design in their analysis, and found no difference in effect size between those studies that were more versus less scientifically rigorous, one lesson that emerged from their work was the importance of interpreting results from meta-analyses within the context of their individual studies. In addition to concerns about the scientific rigor of the primary studies that comprise the meta-analysis, it is essential to consider how well the selected studies represent the population of interest, the treatments studied, and the relevant clinical and functional outcomes.

A number of additional factors should be kept in mind when evaluating the meta-analytic literature. Although this form of analysis has numerous strengths, the process of meta-analysis is not immune to subjectivity or bias (cf. Bangert-Drowns, 1995). For example, one potential source of bias is the "file drawer problem" (Rosenthal, 1979). If clinical research trials with

null results are less likely to be published than other trials, meta-analyses may overestimate treatment effect sizes. One way to address this source of bias is to solicit and include data from unpublished trials in the meta-analysis (see Kirsch, Moore, Scoboria, & Nicholls, 2002, for one example), yet this approach is not always feasible. Another source of bias may be the meta-analyst him- or herself. As elucidated by Bangert-Drowns (1995), the meta-analyst must make decisions about where to search for studies, which studies to include, which outcomes to target, how to code the data, what analytic strategies to employ, and how to interpret the data. All of these phases provide opportunities for "biases to creep into the review process" (Bangert-Drowns, p. 309). Yet in contrast to the narrative review, in which these decisions are largely kept private, one advantage of the meta-analysis is that its procedures are meant to be documented and quantified, so as to inform interpretation by the reader.

Finally, meta-analysis may not always be appropriate (Bangert-Drowns, 1995). Meta-analysis relies on large groups of studies investigating the same question, using similar experimental or quasi-experimental designs. Yet there may be circumstances when such research is unavailable, as in the review of "cutting-edge" developments in a field, early-stage treatment development, integration of different fields of research, or inclusion of all types of studies that evaluate a scientific question. In such circumstances, a narrative review might be most appropriate.

Nevertheless, meta-analysis is a robust approach to evaluation of treatment effects, using well-defined, quantifiable methods. Since the original meta-analysis conducted by Smith et al. (1980), there has been a proliferation of meta-analyses to evaluate psychotherapy outcomes. For example, meta-analyses have been published evaluating the effects of CBT (e.g., Butler et al., 2006), psychodynamic therapy (e.g., Leichsenring et al., 2004), motivational interviewing (e.g., Hettema, Steele, & Miller, 2005), family therapy (e.g., Shadish & Baldwin, 2003), as well as numerous other treatment modalities for a large variety of clinical outcomes.

PLACEBO AND NONSPECIFIC FACTORS

When examining the efficacy and effectiveness of an intervention, the question arises as how best to control for and better examine the contribution of extraneous factors not related to the treatment method specifically. However, the understanding of the "placebo" concept differs in psychosocial versus medical contexts and is described in detail below.

Terminology

The Latin psalm verse "*Placebo Domino in regione vivorum*" ("I shall please the Lord in the land of the living") was sung in the Middle Ages by priests as an intercessory ministration at the deathbed (Walach, 2003).

However, nonecclesiastic agents were often paid to provide surrogacy when the bona fide ministration was not possible or inconvenient. Should the patient recover following the surrogate prayers, the recovery was viewed as a "placebo (I shall please) effect" rather than bona fide ecclesiastic ministration. Thus early on, a disjunct occurred between the specificity of the right prayer being sung and the specificity of the right person singing the right prayer.

In the 18th century, physicians often knew their ministrations (e.g., blood lettings, purgings, potions) were not remedial for the disorder, but administered them in the hope of "pleasing" the patient. The "placebo" construct can still be applied in some aspects of modern medicine when the modality of the procedure (e.g., pill capsule) can be clearly demarcated from the content of the procedure (chemical compound). However, the placebo construct has limited currency where the modality and content are not so clearly demarcated as in psychosocial (i.e., psychotherapeutic) treatments. Instead of the term "placebo," the modality of treatment has been referred to as "nonspecific factors" (Critelli & Neuman, 1984; Shapiro, 1971), which include three broad categories: (1) common factors, (2) factors without specific activity, and (3) unspecified but active factors. *Common factors* are those that are not specific to particular treatments but common to most types of therapy. These include suggestion, persuasion, treatment credibility, therapist attention, and expectancy, demand for improvement, effort justification, therapist allegiance, and therapeutic alliance (Axsom & Cooper, 1985; Evans, 1985; Gaffan, Tsaousis, & Kemp-Wheeler, 1995; Luborsky et al., 1999; Wampold, 2001).

Factors without specific activity are those that underlie panaceas for many different clinical symptoms as they are unlikely to exert general clinical effects through a specific mechanism of action. The several general effects may be the result of several mechanisms of action that derive from a common ingredient. A medical example would be the use of aspirin for any number of somatic problems from headache to prevention of myocardial infarction. That is, the multiple beneficial consequences of aspirin are the result of multiple and general mechanisms.

Unspecified (but active) factors are treatment characteristics that have not been specified as the active ingredients of a particular therapy or treatment. They are not easily derived from the medical metaphor and involve the incidental procedural factors that serve as the means to deliver a broad range of psychosocial interventions. Factors such as a confiding relationship, healing context, plausible problem explanation, and procedure expected to restore health identified by Frank (1961) that are the vehicles and context for healing practices in general are most applicable to this category.

These processes seem to be the analogue of medical placebo and its effects. Psychosocial treatments that allege to be specific interventions must provide additional content that goes beyond nonspecific factors and that can be demarcated from them.

Psychological Analysis of Nonspecific Factors

The construct of "nonspecific factors" has been improved by Grünbaum (1985), who provided a different terminology to compare the concepts of specific and nonspecific treatment factors. An identifiable treatment is predicated on a theory of how that treatment functions and why that treatment is applied to a target disorder. The *characteristic features* of a treatment are those that are required and are based on the theory of action and/or on the theory of the disorder. In contrast, *incidental features* of a treatment are not necessarily derived from the theory and are not required for the treatment procedure.

Furthermore, in Grünbaum's (1985) framework, a treatment can be an unintentional or intentional placebo. In the case of an *unintentional placebo*, the characteristic feature of treatment is not remedial for the disorder, but is only believed to be so by the therapist and the patient. However, the therapist believes the characteristic feature (alleged active ingredient) to be remedial for the problem, and the therapist causes the patient to believe the same. In an *intentional placebo*, the therapist knows that none of the treatment features is remedial. This is the case where an intentional placebo may be administered in a randomized controlled trial to determine whether the characteristic features of an experimental treatment (e.g., drug) provide efficacy above and beyond the incidental features (e.g., placebo capsule) of the treatment application.

In contrast to placebos, bona fide treatments are those in which the specific content of the treatment goes beyond placebo factors or placebo effects. One type is an *inadvertent treatment* that is remedial (i.e., curative or corrective) for the disorder, but where none of the characteristic features of the treatment is actually remedial for the disorder. However, the therapist may believe that all of the characteristic features are remedial for the disorder and that the incidental features are not. Moreover, the therapist may cause the patient to believe that the treatment remediates the disorder because of characteristic features. However, unbeknownst to the therapist, the incidental features of a treatment may actually be responsible for its therapeutic effects. As a contemporary example, component-controlled experimental analyses of eye movement desensitization and reprocessing (EMDR; Davidson & Parker, 2001; F. Shapiro, 2001) have demonstrated that the characteristic eye movement feature is neither a necessary nor a useful component in treatment efficacy. Rather, any measurable benefits of the procedure are likely due to the imagery exposure procedure, which is incidental to the eye movement technique (McNally, 1999).

Intentional treatment occurs where treatment application is remedial for the disorder and where some or all of the characteristic features are remedial for the disorder. In addition, the therapist believes (and may cause the patient to believe) that all of the characteristic features are remedial and that the incidental features are not. Intentional treatment is exemplified by

the circumstance in which the characteristic features do have demonstrable influence and the incidental features might have demonstrable influence. In the case of systematic desensitization for phobia (Wolpe, 1969), the therapist uses relaxation and imagery exposure but correctly attributes efficacy to exposure rather than relaxation.

Finally, *intentional specific treatment* occurs where some or all of the characteristic features are remedial for the disorder, none of the incidental features are remedial for the disorder, these facts are empirically known, and the therapist informs the patient of these facts. Intentional specific treatment is more narrowly defined than intentional treatment. Intentional specific treatment rules out the effects of incidental features and requires the application of characteristic treatment features. Thus, intentional specific treatment gets closer to identifying the theoretical "active ingredients" of the treatment process. In practical terms, this is the case when a therapist is administering the correct procedure for all the right reasons and intentions—that is, when the therapist knows all of the elements of the procedure, but also knows the correct classification of the characteristic and the incidental features. Although this is theoretically possible, intentional specific treatments have been difficult to achieve in psychological intervention research.

The definition of intentional specific treatment provides a more rigorous analysis and demonstration of treatment efficacy and effectiveness. Wait-list and attentional controls are necessary in the analysis of treatment specificity, but they provide for the assessment and manipulation of only the most general, nonspecific effects of treatment (Borkovec & Bauer, 1982). Strong tests require the specification of incidental and characteristic features of treatments. Further, the use of component-control experimental designs is necessary to analyze whether the characteristic features (prescriptive components) are truly "active ingredients" responsible for apparent efficacy, or are instead incidental features that operate as common or nonspecific factors. Treatments should be considered intentional, specific treatments if component-control experimental designs reveal that characteristic components provide a meaningful increment of efficacy beyond that provided by factors that are nonspecific to the treatment. The most relevant experimental controls for strong tests would rely on additive and subtractive (dismantling) designs manipulating incidental and characteristic features (Borkovec, 1985). Doing so will improve the ability to identify separate and incremental contributions to treatment efficacy (Hazlett-Stevens & Borkovec, 1998).

Implications

It is clear that wait-list control comparisons are necessary but not enough for a strong inference regarding whether a specific treatment is effective for a given problem and how that treatment delivers its intended beneficial

effects. We must make a serious effort to clarify the meaning of "placebo" in terms of characteristic features versus incidental features of psychosocial treatments. These definitions need to be based on theories of psychopathology and the theories by which treatments are intended to have their beneficial effects (Eifert, 1996). In this way, we may improve the ability to develop specific treatments that have intended effects on specific aspects of a given disorder. Theory-based treatments must also be compared with meaningful nonspecific factor and component-control conditions. If the content of psychosocial treatments cannot be linked to the theoretical mechanisms of the disorder and therapeutic change, it will be impossible to conduct efficacy research with control conditions that test the incremental validity of treatment specifics.

It is no longer enough to do what one believes at the time to be the right clinical things and hope for the best. Psychosocial treatments should do the right clinical things, and they should be done for right theoretical and empirical reasons. The application of such treatments will meet the challenge of current and future social forces that demand the accountability and cost-effectiveness of psychosocial treatments.

THERAPIST-RELATED FACTORS IN PSYCHOTHERAPY RESEARCH

The Therapeutic/Working Alliance

Therapeutic alliance is among the most widely studied "nonspecific" or "common" factors in psychotherapy. Although definitions for therapeutic alliance have varied, in general it can be described as "the collaborative and affective bond between therapist and patient" (Martin, Garske, & Davis, 2000, p. 438). Three main themes have emerged across the various definitions of alliance: (1) the collaborative nature of the relationship; (2) the affective bond between patient and therapist; and (3) the patient's and therapist's ability to agree on treatment goals (Horvath & Symonds, 1991). Over the past several decades, numerous studies have shown that the strength of the alliance between therapist and patient is related to treatment outcome. Some authors have also posited that the interest in alliance as a predictive factor of treatment outcome has been fostered by a general inability to find significant differences in outcome across different psychotherapeutic approaches (Martin et al., 2000), although other research suggests that differences exist among some psychotherapies (e.g., Tolin, 2010).

As noted earlier, studies have found a significant relationship between alliance and treatment outcome. For example, a study using data from the NIMH TDCRP study (Krupnick et al., 1996) showed that both early and mean alliance ratings were related to treatment outcome, and this relationship did not differ across the treatment modalities. Recent meta-analyses have confirmed a significant alliance–outcome relationship, with correlations typically around $r = .20$, which is small to medium in magnitude (Del

Re, Flückiger, Horvath, Symonds, & Wampold, 2012; Martin et al., 2000). Of particular interest to researchers is the degree to which the therapist and/or patient contributes to the alliance, which subsequently has an effect on outcome. Krupnick et al. (1996) found that patient, but not therapist, contribution to alliance was related to outcome (accounting for 8% of the variance in outcomes with early alliance and 21% of the variance with mean alliance scores). More recent examinations of the TDCRP study, however, have obtained different results. Zuroff, Kelly, Leybman, Blatt, and Wampold (2010) reanalyzed data from 157 patients in the TDCRP study, finding that between-therapist variability in alliance was significantly related to outcome, regardless of the severity of the initial depression or the type of treatment used (CBT vs. interpersonal therapy vs. placebo plus clinical management). They also found a significant within-therapist effect (i.e., patient contribution to alliance), but to a lesser degree.

In contrast, other studies have indicated that the therapist contribution to alliance is more significant than the patient contribution. For example, a naturalistic study of a heterogeneous sample of counseling center patients from the Research Consortium of Counseling and Psychological Services in Higher Education study (Baldwin, Wampold, & Imel, 2007) found that patients who rated alliance with their therapist highly experienced better outcomes compared with patients who rated the alliance with their therapist lower (i.e., between-therapist effects or therapist contribution to alliance). In contrast, patient variability in alliance was not related to outcome. In other words, the treatment outcome of patients who were able to form a good alliance with their therapist did not differ from that of patients who rated poor alliance with that same therapist. A recent meta-analysis by Del Re et al. (2012) also suggested that therapist effects were more closely associated with the alliance–outcome relationship than were patient effects. For instance, they found that the between-therapist alliance–outcome correlation was $r = 40$, even after controlling for other potential moderating variables (e.g., type of treatment or outcome measure used). Based on their analyses, they concluded that therapist actions and characteristics were more influential in the alliance-outcome relationship compared with patient characteristics, although they also noted that their study did not rule out the possibility of an interaction between therapist and patient actions/characteristics.

Efforts have been made to examine other potential moderators of the alliance–outcome relationship. Based on prior results suggesting that therapist actions and characteristics may have more of an effect on alliance and thus outcome, Ulvenes et al. (2012) examined the relationship between therapist actions, bond, and outcomes between 50 patients with Cluster C (anxious, fearful) personality disorders who received either cognitive therapy (CT) or short-term dynamic psychotherapy (STDP). Consistent with prior studies, across the total sample, the therapist–patient bond was positively related to outcome. Interestingly, the authors also

found that the more that the therapist avoided eliciting or commenting on affective content in session, the more highly the patient rated the bond with his/her therapist; yet, this aspect of bond was unrelated to outcome. A glimpse within treatment modalities yielded further interesting results. Within STDP, a therapist's focus on affect was related to outcome even if the patient's rating on the therapeutic bond decreased. In other words, a therapist's focus on affect decreased the perceived bond between the therapist and the patient but improved outcomes. Within CT, a different picture emerged: Therapist avoidance of affect was positively related to therapeutic bond and symptom reduction. Although this study suggests that alliance differs by treatment modality, results from a recent meta-analysis found contradictory results showing that type of treatment, along with type of research design, use of disorder-specific treatment manuals, and use of particular outcome measures were not significant moderators of the alliance–outcome relationship (Flückiger, Del Re, Wampold, Symonds, & Horvath, 2012).

The emphasis on alliance in predicting outcome has not been without its critics. For example, some have argued that prior studies have not been conducted in a way that would determine the direction of the relationship: Good outcomes could result in good alliance, rather than vice versa (DeRubeis, Brotman, & Gibbons, 2005). In fact, some studies have shown that later alliance was predicted by early improvement, but later improvement was not predicted by early alliance (Feeley, DeRubeis, & Gelfand, 1999; Tang & DeRubeis, 1999). However, some studies have found that early alliance predicted later outcomes for the cognitive-behavioral analysis system of psychotherapy (CBASP; Klein et al., 2003) and brief dynamic therapy (Barber, Connolly, Crits-Christoph, Gladis, & Siqueland, 2000). Few, if any, studies have experimentally manipulated alliance, and some have argued that it is unethical to do so (Wampold, 2005), thus limiting what can be known about the alliance–outcome relationship. Therefore, the debate continues as to the nature and degree of the alliance.

Therapist Effects

In addition to the relationship between the therapist and client, what role do therapist characteristics play in treatment outcome? This question carries important implications for practice, research, and training. As suggested by Hill (2006), if it is the treatment that accounts for most of the variance in outcome, then it is important that we have detailed treatment manuals that specify what any therapist should do within the therapy process. Yet if it is the therapist who is largely responsible for the outcome, then it is important to select and train therapists to maximize the characteristics that make them effective. Of course, there may also be client characteristics that drive treatment response, or some interaction of treatment, therapist, and client factors that best predict outcome. In sum, identification of the sources of

treatment effects has the potential to elucidate Gordon Paul's famous 1967 question.

Yet the question of "what treatment?" has been the focus of much greater scientific scrutiny than the question of "by whom?" Indeed, the question is much more difficult to answer in any one research study, given the typically small number of therapists who are enlisted to deliver interventions in clinical trials, and the efforts made by investigators to minimize therapist heterogeneity (e.g., through careful selection, frequent and intensive supervision, manualization of therapy protocols, and therapist fidelity checks). Such efforts are important if one is most interested in evaluating the efficacy of a particular treatment, as they contribute to the maximization of internal validity. Yet they result in datasets that are not as well suited for answering questions about therapist effects. Thus, what we do know about therapist effects has been derived largely from meta-analyses and other secondary analyses of clinical trials data.

In an early investigation of therapist effects, Luborsky et al. (1986) reanalyzed data from four psychotherapy outcome trials and found evidence for differences between therapists in clinical outcome. Further, they concluded that therapist differences were stronger predictors of outcome than the treatments themselves. Shortly thereafter, Crits-Christoph and colleagues (1991) published a meta-analysis that revealed great variability in therapist effects across 27 studies; on average, 8.6% of the variance in outcome could be accounted for by the therapists themselves, translating into a medium effect size. Support for significant therapist effects on psychotherapy outcome has been replicated across several samples (e.g., Luborsky, McLellan, Diguer, Woody, & Seligman, 1997; Huppert et al., 2001), including a large study (N = 1,779) of clinical outcomes in a college counseling center and in secondary analyses of data from a number of prominent clinical trials, including Project MATCH (Project MATCH Research Group, 1998), the NIDA Cocaine Collaborative Study (Siqueland et al., 2000), and the TDCRP (Blatt, Sanislow, Zuroff, & Pilkonis, 1996; Kim, Wampold, & Bolt, 2006).

Despite several consistencies across these studies, therapist effects in psychotherapy outcome remain controversial (for a review, see Crits-Christoph & Gallop, 2006). This controversy is perhaps best exemplified in the TDCRP study (Elkin et al., 1989), in which Kim et al. (2006) estimated that 8% of the variance in outcome was due to therapist effects, whereas 0% was accounted for by the treatments. Yet in a separate analysis by Elkin et al. (2006) therapist effects were found to be essentially null. Several explanations for these disparate results have been posited, including differences in samples (e.g., completers alone vs. intent to treat, that is, analyses including both completers and dropouts), in outcome measures selected for analysis, and in the statistical models employed (Crits-Christoph & Gallop, 2006; Soldz, 2006; Wampold & Bolt, 2006). Crits-Christoph and Gallop (2006) also concluded that any therapist effects within the TDCRP may

have been largely explained by a study design that relied on comparison of two active treatments with generally equal efficacy, in the absence of a control condition. Thus, even if differences between therapists explained more variance in the outcome than the treatments themselves, this result does not necessarily mean that the treatments were unimportant.

Nevertheless, there remains evidence that therapist effects do predict clinical outcomes following psychotherapy. Attempts to elucidate specific therapist characteristics that influence outcome, however, have been somewhat mixed. Within the TDCRP, for example, Blatt et al. (1996) reported that more effective psychotherapists had a more psychological (vs. biological) orientation and the expectation that effective therapy would require more sessions. These investigators did not find differences in therapeutic efficacy among therapists with more versus less clinical experience, which has been further supported by data derived from a number of other clinical trials (Beutler, 1997). Factors such as type of professional training (Christensen & Jacobson, 1994) and theoretical orientation (Garske & Anderson, 2003) have also failed to distinguish effective from less effective therapists. If largely unexplained, some have argued that evidence for therapist effects implies a greater need for well-described, manualized interventions. Indeed, in their meta-analysis of the relevant literature, Crits-Christoph et al. (1991) found a diminished influence of therapist effects on outcome when treatment manuals were employed.

Yet others have argued that it would be premature to abandon research into the characteristics that account for therapist effects in psychotherapy research. For example, research has suggested that effective therapists are more likely to demonstrate warmth and positive communication behaviors in their treatment sessions (Henry, Schacht, Strupp, Butler, & Binder, 1993; Westerman, Foote, & Winston, 1995), to have a more complex understanding of clinical material (Goldfried, Raue, & Castonguay, 1998), and to be more successful in establishing a positive therapeutic alliance with their patients (Luborsky et al., 1997), although replication of these findings will be necessary.

EVIDENCE-BASED PRACTICE IN PSYCHOLOGY

It frequently has been difficult to increase the adoption of research-based psychotherapies by therapists in routine practice (Becker, Zeyfert, & Anderson, 2004; Hoagwood, Burnas, Kiser, Ringeisen, & Schoenwald, 2001). In medicine, physicians are more consistently held to the standard of evidence-based practice, which is defined as "the conscientious and judicious use of current best evidence from clinical care research in the management of individual patients" (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, pp. 71–72). In practice, this requires the integration of overlapping "spheres" of knowledge in decision making that include research evidence,

clinical experience, and patient preferences (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000). Since their formal adoption in medicine, various related health disciplines, such as nursing and social work, have formulated similar evidence-based practice proposals (Satterfield et al., 2009). However, the road to evidence-based practice has been long and circuitous in psychology.

In 1993, the American Psychological Association's Division of Clinical Psychology entered this area by forming a task force to define empirically validated treatments, later termed "empirically supported treatments" (ESTs) (Chambless & Ollendick, 2001). The task force issued its first report in 1995 (Task Force, 1995), with subsequent updates in 1996 and 1998 (Chambless et al., 1996, 1998). The task force developed decision rules for defining different levels of support for ESTs and formulated lists of psychotherapies meeting these criteria. Designation as a "well-established treatment" requires two randomized trials demonstrating efficacy compared with a "placebo" or another established treatment, or a large series of single-case design experiments showing the same. These trials must also include the use of treatment manuals, specify the sample clearly, include samples of adequate size, and be conducted by independent investigator teams to minimize bias. In contrast, "probably efficacious treatments" must show evidence of efficacy under controlled conditions but do not fully meet criteria for well-established treatments. Recently, a new website was launched that provides an updated listing of ESTs (*www.psychologicaltreatments.org*), and also designates "controversial treatments" as those yielding conflicting results or making unsupported claims about their mechanisms of action (e.g., EMDR).

One potential benefit of defining ESTs is its utility in helping clinicians and the public differentiate between science-based versus pseudoscientific treatments. Establishing well-defined criteria for evaluating the merits of different types of treatments has identified those treatments that currently have the strongest empirical support and tend to adhere to the principles of science in their testing. However, the EST criteria are sufficiently broad that treatments that have characteristics of pseudoscience can still be considered "empirically supported" under current guidelines. For example, Herbert, Lilienfeld, Lohr, Montgomery, O'Donohue, et al. (2000) noted that EMDR, a treatment for trauma that is claimed to work through "bilateral stimulation" (e.g., alternating eye movements), quickly began to show up on EST lists, even though its mechanisms of action remain controversial and its promotion has been characterized by exaggerated claims of effectiveness that go well beyond the data. More recently, proponents of so-called "energy" therapies, in which clients are instructed to tap on their bodies in prescribed sequences in a form of "psychological acupuncture," have claimed that this type of intervention also meets the requirements to be listed as an EST based on the current criteria (Feinstein, 2008). Others have noted the many pseudoscientific characteristics

of energy therapies and have questioned the quality of evidence provided to support them (Devilly, 2005; Pignotti & Thyer, 2009b). Thus, although EST criteria have been helpful in identifying scientifically supported treatments, they are not foolproof; further evaluation of a treatment's underlying theory and dissemination practices is warranted before uncritically accepting it into practice.

Furthermore, EST lists can only be considered one part of the larger process of evidence-based practice, as the task force's product provided no specific guidance about how this information should be integrated with other knowledge (e.g., emerging research, idiographic assessment) in the clinical decision-making process, which is the hallmark of evidence-based practice (Spring, 2007). Lists of ESTs were intended to represent a first step toward promoting the evidence-based practice of psychotherapy. Psychology EST lists have occurred in the context of a growing movement over recent decades promoting mental health practices based on the best available scientific evidence. In the United Kingdom, the National Institute for Health and Clinical Excellence (*www.nice.org.uk*) publishes treatment guidelines in medicine and psychiatry. In the United States, the National Registry of Evidence-Based Programs and Practices (NREPP) similarly lists evidence-based treatments and is sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA; *www.nrepp.samhsa.gov*).

Some clinicians question the usefulness of the application of scientific methods in general, and RCTs in particular, in guiding the practice of psychotherapy. They also object to the apparent loss of clinician autonomy that the evidence-based practice movement can entail (Levant, 2004). For example, it has been argued that the use of specific treatment manuals in a prescriptive manner reduces clinical judgment and the ability to apply a treatment flexibly, thus limiting the usefulness of the treatment in clinical practice (Westen, Novotny, & Thompson-Brenner, 2004). Scientist–practitioners tend not to object to efforts to promote empirically supported practice in principle. Nevertheless, Herbert (2003) outlined several problems, including the inadequacy of wait-list control conditions, for determining EST status. That is, because most psychotherapies can be shown to be superior to no treatment, simple comparisons to no-treatment controls are of little practical value in defining ESTs (beyond establishing initial safety). In addition, the RCTs often used to establish EST status have been criticized for including clients who are not representative of those seen in routine clinical practice (e.g., they are less diagnostically complex compared to clients in naturalistic settings; Westen et al., 2004). Furthermore, determining that a treatment works does not necessarily tell us why a treatment works. Rosen and Davison (2003) proposed that efforts should be made to identify empirically supported *principles* of change, rather than proprietary therapies. However, the task of identifying principles of change is even more complicated in practice because clinicians from different theoretical orientations

often disagree as to what these core principles are and how they should be defined.

It was not until 2006 that a new APA Presidential Task Force published its formal conceptualization of evidence-based practice in psychology. Similar to the definition originally proposed by Sackett et al. (1996), the task force defined evidence-based practice in psychology as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (APA Presidential Task Force on Evidence-Based Practice, 2006, p. 273). Some have criticized the American Psychological Association's definition for placing too much weight on the idea of clinical expertise and intuition rather than on research evidence when making treatment decisions (Stuart & Lilienfeld, 2007). Still, the task force's document represents a positive step forward in promoting evidence-based practice, especially considering that psychology traditionally has lagged behind related health disciplines in this area (Herbert & Gaudiano, 2005). Efforts also are currently underway to develop more formal treatment guidelines in psychology (Kurtzman & Bufka, 2011). A treatment guideline is essentially a blueprint for treating a particular condition. Ideally, treatment guidelines specify an algorithm of decision rules based on the best available scientific data specifying recommended frontline and secondary interventions.

CONCLUSION

The application of scientific methods to the development, testing, and promotion of psychotherapies has produced numerous effective interventions for individuals suffering from psychological disorders and their families. Research has demonstrated that many of these interventions are as or more effective than psychiatric medications for common conditions, such as mood and anxiety disorders. Furthermore, treatment guidelines often recommend evidence-based psychotherapies as frontline approaches for children, adults, and the elderly. However, many practicing therapists still are using untested or less effective approaches, and more work is needed to promote the use of evidence-based practices. Modern psychotherapy research has evolved over the years and has become increasingly sophisticated. Recent developments have focused more on producing practical, feasible, and flexible therapies that better meet the needs of practicing clinicians, which should continue to improve their uptake in the community. More sophisticated statistical techniques have fostered increased attention to elucidating the "active ingredients" of effective treatments so that they can be improved and refined. For psychotherapy to continue to progress, it will be essential that we continue to test our interventions to verify their safety and efficacy and to ensure that they are grounded in the latest developments from basic psychological science.

GLOSSARY

Dodo bird verdict: The oft-contested conclusion that all credible psychotherapies work equally well based on meta-analyses that show similar effects across treatments.

Effectiveness: The effects of a treatment when tested in "real-world" settings or under typical clinical conditions.

Efficacy: The effects of a treatment when tested in controlled research so that cause-and-effect conclusions can be drawn.

Empirically supported treatments: Particular treatments that have been shown to be efficacious for certain conditions based on results from controlled research.

Evidence-based practice: Using the current best evidence available from research (e.g., randomized controlled trials) when making clinical decisions. Includes consideration and integration of relevant research evidence with clinical, patient, and other contextual factors.

Mediator: An intervening variable that statistically accounts for the effect of an independent variable on a dependent variable. Used to investigate the potential mechanisms of change or processes through which a treatment achieves its effect.

Meta-analysis: A set of quantitative procedures by which effect sizes, or statistical calculations representing the magnitude of treatment effects, are derived from multiple studies and then combined into a common effect size for interpretation purposes.

Moderator: An individual (e.g., symptom severity) or contextual (e.g., gender) variable that influences whether a treatment is more or less likely to produce an effect.

Nonspecific factors: The common (i.e., found across most treatments), general (i.e., working through nonspecific mechanisms), or unspecified (i.e., derived from incidental procedures or the treatment context itself) features of treatments that can sometimes produce beneficial effects. Also, sometimes referred to as "placebo."

REFERENCES

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68*, 331–339.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271–285.

Axsom, D., & Cooper, J. (1985). Cognitive dissonance and psychotherapy: The role of effort justification in inducing weight loss. *Journal of Experimental Social Psychology, 21*, 149–160.

Baardseth, T. P., Goldberg, S. B., Pace, B. T., Wislocki, A. P., Frost, N. D., Siddiqui, J. R., et al. (2013). Cognitive-behavioral therapy versus other therapies: Redux. *Clinical Psychology Review, 33*, 395–405.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75*, 842–852.

Bangert-Drowns, R. L. (1995). Misunderstanding meta-analysis. *Evaluation and the Health Professions, 18*, 304–314.

Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. *Journal of Consulting and Clinical Psychology, 68,* 1027–1032.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Becker, C. B., Zeyfert, C., & Anderson, E. (2004). A survey of psychologists' attitudes towards and utilization of exposure therapy for PTSD. *Behavior Research and Therapy, 42*, 277–293.

Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change*. New York: Wiley.

Beutler, L. E. (1997). The psychotherapist as a neglected variable in psychotherapy: An illustration by reference to the role of therapist experience and training. *Clinical Psychology–Science and Practice, 4*, 44–52.

Blatt, S. J., Zuroff, D. C., Bondi, C. M., & Sanislow, C. A. (2000). Short- and long-term effects of medication and psychotherapy in the brief treatment of depression: Further analyses of data from the NIMH TDCRP. *Psychotherapy Research, 10,* 215–234.

Blatt, S. J., Sanislow, C. A., III, Zuroff, D. C., & Pilkonis, P. A. (1996). Characteristics of effective therapists: Further analyses of data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*, 1276–1284.

Borkovec, T. D. (1985). Placebo: Defining the unknown. In L. White, B. Tursky, & G. E. Schwartz (Eds.), *Placebo: Theory, research, and mechanisms* (pp. 59–64). New York: Guilford Press.

Borkovec, T. D., & Bauer, R. (1982). Experimental design in group outcome research. In A. Bellack, M. Hersen, & A. Kazdin (Eds.), *International handbook of behavior modification and therapy* (pp. 139–165). New York: Plenum Press.

Borkovec, T. D., Echemendia, R. J., Ragusea, S. A., & Ruiz, M. (2001). The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice, 8,* 155–167.

Budd, R., & Hughes, I. (2009). The Dodo bird verdict—controversial, inevitable and important: A commentary on 30 years of meta-analyses. *Clinical Psychology and Psychotherapy, 16,* 510–522.

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006).The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review, 26*, 17–31.

Chambless, D., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P., et al. (1998). Update on empirically validated therapies: II. *Clinical Psychologist, 51*, 3–16.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685–716.

Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *Clinical Psychologist, 49*, 5–18.

Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy—the status and challenge of nonprofessional therapies. *Psychological Science, 5*, 8–14.

Cioffi, F. (1998). *Freud and the question of pseudoscience*. Chicago: Open Court.

Clarke, G. N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology, 63,* 718–725.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Critelli, J. W., & Neumann, K. F. (1984). The placebo: Conceptual analysis of a construct in transition. *American Psychologist, 39*, 32–39.

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., et al. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1*, 81–91.

Crits-Christoph, P., & Gallop, R. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program and other psychotherapy studies. *Psychotherapy Research, 16*, 178–181.

Davidson, P. R., & Parker, C. H. (2001). Eye movement desensitization and reprocessing (EMDR): A meta-analysis. *Journal of Consulting and Clinical Psychology, 69*, 305–316.

Del Re, A. C., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance–outcome relationship: A restricted-maximum likelihood meta-analysis. *Clinical Psychology Review, 32,* 642–649.

DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice, 12,* 174–183.

Devilly, G. J. (2005). Power therapies and possible threats to the science of psychology and psychiatry. *Australian and New Zealand Journal of Psychiatry, 39*, 437–445.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., et al. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*, 658–670.

Eifert, G. H. (1996). More theory-driven and less diagnosis-based behavior therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 27*, 75–86.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. T., Sotsky, S. M., Collins, J. F., et al. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971–982.

Elkin, I., Falconnier, L., Martinovich, Z., & Mahoney, C. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Psychotherapy Research, 16*, 144–160.

Evans, F. J. (1985). Expectancy, therapeutic instructions, and the placebo response. In L. White, B. Tursky, & G. E. Schwartz (Eds.), *Placebo: Theory, research, and mechanisms* (pp. 215–228). New York: Guilford Press.

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16,* 319–324.

Eysenck, H. J. (1961). The effects of psychotherapy. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology* (pp. 697–725). New York: Basic Books.

Eysenck, H. J. (1966). *The effects of psychotherapy*. New York: International Science Press.

Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 67,* 578–582.

Feinstein, D. (2008). Energy psychology: A review of the preliminary evidence. *Psychotherapy Theory, Research, Practice, Training, 45*, 199–213.

Flückiger, C., Del Re, A. C., Wampold, B. E., Symonds, D., & Horvath, A. O. (2012). How central is the alliance in psychotherapy?: A multilevel longitudinal meta-analysis. *Journal of Counseling Psychology, 59,* 10–17.

Frank, G. (1961). *Persuasion and healing*. Baltimore: Johns Hopkins University Press.

Freiheit, S. R., Vye, C., Swan, R., & Cady, M. (2004). Cognitive-behavioral therapy for anxiety: Is dissemination working? *Behavior Therapist, 27*, 25–32.

Gaffan, E. A., Tsaousis, I., & Kemp-Wheeler, S. M. (1995). Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 63*, 966–980.

Garske, J., & Anderson, T. (2003). Toward a science of psychotherapy research: Present status and evaluation. In S. Lilienfeld, S. Lynn, & J. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 17–38). New York: Guilford Press.

Gaudiano, B. A., Brown, L. A., & Miller, I. W. (2012). Let your intuition be your guide?: Individual differences in the evidence-based practice attitudes of psychotherapists. *Journal of Evaluation in Clinical Practice, 17*, 628–634.

Gaudiano, B. A., Brown, L. A., & Miller, I. W. (2011). Tapping their patients' problems away?: Characteristics of psychotherapists using energy meridian techniques. *Research on Social Work Practice, 22*, 647–655.

Gaudiano, B. A., & Herbert, J. D. (2000). Can we really tap away our problems?: A critical analysis of thought field therapy. *Skeptical Inquirer, 24*, 29–33, 36.

Glass, G. V. (1976). Primary, secondary, and meta-analysis research. *Educational Researcher, 5*, 3–8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Goldfried, M. R., Raue, P. J., & Castonguay, L. G. (1998). The therapeutic focus in significant sessions of master therapists: A comparison of cognitive-behavioral and psychodynamic–interpersonal interventions. *Journal of Consulting and Clinical Psychology, 66*, 803–810.

Grissom, R. J. (1996). The magical number .7 ± .2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *Journal of Consulting and Clinical Psychology, 64*, 973–982.

Grünbaum, A. (1985). Explication and implications of the placebo concept. In L. White, B. Tursky, & G. E. Schwartz, (Eds.), *Placebo: Theory, research, and mechanisms* (pp. 9–36). New York: Guilford Press.

Hayes, S. C., Long, D. M., Levin, M. E., & Follette, W. C. (2013). Treatment development: Can we find a better way? *Clinical Psychology Review*, *33*(7), 870–882.

Hazlett-Stevens, H., & Borkovec, T. D. (1998). Experimental design and

methodology in between-group intervention outcome research. In R. Schulz, M. P. Lawton, & G. Maddox (Eds.), *Annual review of gerontology and geriatrics: Vol. 18. Intervention research with older adults* (pp. 17–47). New York: Springer.

Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy—mediators of therapists responses to training. *Journal of Consulting and Clinical Psychology, 61*, 441–447.

Herbert, J. D. (2003). The science and practice of empirically supported treatments. *Behavior Modification, 27*, 412–430.

Herbert, J. D., & Gaudiano, B. A. (2005). Moving from empirically supported treatment lists to practice guidelines in psychotherapy: The role of the placebo concept. *Journal of Clinical Psychology, 61*, 893–908.

Herbert, J. D., Lilienfeld, S. O., Lohr, J. M., Montgomery, R. W., O'Donohue, W. T., Rosen, G. M., et al. (2000). Science and pseudoscience in the development of eye movement desensitization and reprocessing: Implications for clinical psychology. *Clinical Psychology Review, 20*, 945–971.

Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annual Review of Clinical Psychology, 1*, 91–111.

Hill, C. E. (2006). Introduction to special section on therapist effects. *Psychotherapy Research, 16*, 143.

Hoagwood, K., Burns, B. J., Kiser, L., Ringeisen, H., & Schoenwald, S. K. (2001). Evidence-based practice in child and adolescent mental health services. *Psychiatric Services, 52*, 1179–1189.

Hopko, D. R., Lejuez, C. W., Ruggiero, K. J., & Eifert, G. H. (2003). Contemporary behavioral activation treatments for depression: Procedures, principles, and progress. *Clinical Psychology Review, 23*, 699–717.

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*, 139–149.

Hunsley, J., & Di Giulio, G. (2002). Dodo bird, phoenix, or urban legend?: The question of psychotherapy equivalence. *Scientific Review of Mental Health Practice, 1*, 11–22.

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*, 747–755.

Iacoviello, B. M., McCarthy, K. S., Barrett, M. S., Rynn, M., Gallop, R., & Barber, J. P. (2007). Treatment preferences affect the therapeutic alliance: Implications for randomized controlled trials. *Journal of Consulting and Clinical Psychology, 75*, 194–198.

Imel, Z., & Wampold, B. (2008). The importance of treatment and the science of common factors in psychotherapy. *Handbook of counseling psychology* (4th ed., pp. 249–262). Hoboken, NJ: Wiley.

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., et al. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64*, 295–304.

Jacobson, N. S., & Hollon, S. D. (1996). Cognitive-behavior therapy versus pharmacotherapy: Now that the jury's returned its verdict, it's time to present

the rest of the evidence. *Journal of Consulting and Clinical Psychology, 64,* 74–80.

Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3,* 1–27.

Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research, 16*, 161–172.

Kirsch, I., Moore, T. J., Scoboria, A., & Nicholls, S. S. (2002). The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention and Treatment, 5*(1), Article 23.

Klein, D. F., & Ross, D. C. (1993). Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program general effectiveness report. *Neuropsychopharmacology, 8,* 241–251.

Klein, D. N., Schwartz, J. E., Santiago, N. J., Vocisano, C., Castonguay, L. G., Arnow, B., et al. (2003). Therapeutic alliance in depression treatment: Controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology, 71,* 997–1006.

Kraemer, H. C., Kiernan, M., Essex, M., & Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the Baron and Kenny and MacArthur approaches. *Health Psychology, 27,* S101–S108.

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry, 63*, 484–489.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*, 877–883.

Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J., et al. (1996). The role of the therapeutic alliance in psychotherapy pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64,* 532–539.

Kurtzman, H., & Bufka, L. (2011). APA moves forward on developing clinical treatment guidelines. *Psychological Science Agenda, 35*(7). Retrieved November 29, 2012, from *www.apa.org/science/about/psa/2011/07/clinical-treatment.aspx*.

Leichsenring, F., Rabung, S., & Leibing, E. (2004). The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders: A meta-analysis. *Archives of General Psychiatry, 61*, 1208–1216.

Levant, R. F. (2004). The empirically validated treatments movement: A practitioner/educator perspective. *Clinical Psychology: Science and Practice, 11*, 219–224.

Leykin, Y., DeRubeis, R. J., Gallop, R., Amsterdam, J. D., Shelton, R. C., & Hollon, S. D. (2007). The relation of patients' treatment preferences to outcome in a randomized clinical trial. *Behavior Therapy, 38,* 209–217.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53–70.

Luborsky, L., Crits-Christoph, P., Mclellan, A. T., Woody, G., Piper, W., Liberman, B., et al. (1986). Do therapists vary much in their success—findings from 4 outcome studies. *American Journal of Orthopsychiatry, 56*, 501–512.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95–106.

Luborsky, L., McLellan, A. T., Diguer, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice, 4*, 53–65.

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "everyone has won and all must have prizes"? *Archives of General Psychiatry, 32,* 995–1008.

Mansell, W. (2011). Core processes of psychopathology and recovery: "Does the Dodo bird effect have wings?" *Clinical Psychology Review, 31,* 189–192.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68,* 438–450.

McEvoy, P. M., & Nathan, P. (2007). Effectiveness of cognitive behavior therapy for diagnostically heterogeneous groups: A benchmarking study. *Journal of Consulting and Clinical Psychology, 75,* 344–350.

McNally, R. J. (1999). On eye movements and animal magnetism: A reply to Greenwald's defense of EMDR. *Journal of Anxiety Disorders, 13*, 617–620.

Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin, 126,* 964–981.

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology and Psychotherapy, 10*, 361–373.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31*, 109–118.

Pignotti, M., & Thyer, B. A. (2009a). Use of novel unsupported and empirically supported therapies by licensed clinical social workers: An exploratory study. *Social Work Research, 33*, 5–17.

Pignotti, M., & Thyer, B. (2009b). Some comments on "Energy psychology: A review of the evidence": Premature conclusions based on incomplete evidence? *Psychotherapy Theory, Research, Practice, Training, 46*, 257–261.

Project MATCH Research Group. (1997). Matching alcoholism treatment to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol, 58*, 7–29.

Project MATCH Research Group. (1998). Therapist effects in three treatments for alcohol problems. *Psychotherapy Research, 8*, 455–474.

Rachman, S., & Wilson, G. T. (1980). *The effects of psychological therapy.* Oxford, UK: Pergamon Press.

Reichenbach, H. (1938). *Experience and prediction: An analysis of the foundations and the structure of knowledge.* Chicago: University of Chicago Press.

Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification, 27*, 300–312.

Rosenthal, R. (1979). The "file drawer" problem and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6*, 412–415.

Roth, A., & Fonagy, P. (1996). *What works for whom: A critical review of psychotherapy research.* New York: Guilford Press.

Rounsaville, B. J, Carroll, K. M., & Onken, L. S. (2001). A stage model of behavioral therapies research: Getting started and moving on from Stage I. *Clinical Psychology: Science and Practice, 8*, 133–142.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*, 71–72.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W. M., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). Philadelphia: Churchill Livingstone.

Satterfield, J. M., Spring, B., Brownson, R. C., Mullen, E. J., Newhouse, R. P., Walker, B. B., et al. (2009). Toward a transdisciplinary model of evidence-based practice. *Milbank Quarterly, 87*, 368–390.

Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: What matters when? *Psychiatric Services, 52*, 1190–1197.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study. *American Psychologist, 50*, 965–974.

Shadish, W. R., & Baldwin, S. A. (2003). Meta-analysis of MFT interventions. *Journal of Marital and Family Therapy, 29*, 547–570.

Shapiro, A. K. (1971). Placebo effects in medicine, psychotherapy, and psychoanalysis. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis* (pp. 439–473). New York: Wiley.

Shapiro, F. (2001). *Eye movement desensitization and reprocessing (EMDR): Basic principles, protocols, and procedures* (2nd ed.). New York: Guilford Press.

Singer, M. T., & Lalich, J. (1996). "*Crazy" therapies*. San Francisco: Jossey-Bass.

Siqueland, L., Crits-Christoph, P., Barber, J. P., Butler, S. F., Thase, M., Najavits, L., et al. (2000). The role of therapist characteristics in training effects in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *Journal of Psychotherapy Practice and Research, 9*, 123–130.

Sloane, R. B., Staples, F. R., Cristol, A. H., Yorkston, N. J., & Whipple, K. (1975). Short-term analytically oriented psychotherapy versus behavior therapy. *American Journal of Psychiatry, 132*, 373–377.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752–760.

Smith, M. L., Glass, G. V., & Miller, T. L. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press.

Soldz, S. (2006). Models and meanings: Therapist effects and the stories we tell. *Psychotherapy Research, 16*, 173–177.

Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. *Journal of Clinical Psychology, 63*, 611–631.

Stewart, R. E., & Chambless, D. L. (2007). Does psychotherapy research inform treatment decisions in private practice? *Journal of Clinical Psychology, 63*, 267–281.

Stuart, R. B., & Lilienfeld, S. O. (2007). The evidence missing from evidence-based practice. *American Psychologist, 62*, 615–616.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67,* 894–904.

Task Force on the Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated treatments. *Clinical Psychologist, 48*, 3–23.

Tolin, D. F. (2010). Is cognitive-behavioral therapy more effective than other therapies?: A meta-analytic review. *Clinical Psychology Review, 30,* 710–720.

Ulvenes, P. G., Berggraf, L., Hoffart, A., Stiles, T. C., Svartberg, M., McCullough, L., et al. (2012). Different processes for different therapies: Therapist actions, therapeutic bond, and outcome. *Psychotherapy, 49,* 291–302.

Walach, H. (2003). Placebo and placebo effects: A concise review. *Focus on Complementary Therapies, 8*, 178–187.

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.

Wampold, B. E. (2005). Establishing specificity in psychotherapy scientifically: Design and evidence issues. *Clinical Psychology: Science and Practice, 12,* 194–197.

Wampold, B. E., & Bolt, D. M. (2006). Therapist effects: Clever ways to make them (and everything else) disappear. *Psychotherapy Research, 16*, 184–187.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin, 122,* 203–215.

Wells, K. B. (1999). Treatment research at the crossroads: The scientific interface of clinical trials and effectiveness research. *American Journal of Psychiatry, 156,* 5–10.

Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy, 43,* 1243–1261.

Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69,* 875–899.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130,* 631–663.

Westerman, M. A., Foote, J. P., & Winston, A. (1995). Change in coordination across phases of psychotherapy and outcome—2 mechanisms for the role played by patients contribution to the alliance. *Journal of Consulting and Clinical Psychology, 63*, 672–675.

Wolpe, J. (1969). *The practice of behavioral therapy.* New York: Pergamon Press.

Zuroff, D. C., Kelly, A. C., Leybman, M. J., Blatt, S. J., & Wampold, B. E. (2010). Between-therapist and within-therapist differences in the quality of the therapeutic relationship: Effects on maladjustment and self-critical perfectionism. *Journal of Clinical Psychology, 66,* 681–697.

CHAPTER SEVEN

New Age and Related Novel Unsupported Therapies in Mental Health Practice

Monica Pignotti and Bruce A. Thyer

Over 10 years have elapsed since the publication of the chapter on this topic in the first edition of this book (Singer & Nievod, 2003). In crafting the second edition of this chapter, we have reflected on the question of how the promotion and use of New Age type therapies by mental health professionals have changed over the past decade. As the promotion on the Internet of these therapies, accompanied by unsupported claims, appears to be more aggressive than ever, some of these therapies have worked their way into mainstream mental health practice, seminars, and journals, resulting in disturbing trends that we will be discussing. Although 10 years ago such therapies were practiced by those on the fringes of the profession, for some of these therapies this appears no longer to be the case.

Now more than ever, many New Age therapies can be classified as pseudoscientific, because they often take on the superficial appearance of legitimate science, when they are anything but. There is little or no more evidence for their claimed theoretical basis (e.g., the body's "energy system") than there ever was; yet some therapies based on unsupported theories are claiming to be research-based and to meet the standards of Division 12 (Society of Clinical Psychology) of the American Psychological Association for a research-supported treatment (Feinstein, 2008, 2012).

In this chapter, we continue the conversation begun by Singer and Nievod (2003) on professional accountability with regard to these types of therapies, as a basis for their evaluation. Recovered memory therapies

(RMTs) and the overdiagnosis of dissociative identify disorder (DID) were discussed in the previous edition of this chapter (see also Lilienfeld & Lynn, Chapter 5, and Lynn, Krackow, Loftus, Locke, & Lilienfeld, Chapter 8, this volume), and we will update these topics. Although the RMTs and their proponents underwent crushing legal defeats during the 1990s, DID remains in the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). Such therapies, though typically couched in more cautious language by their proponents, continue in various forms including "parts work," a type of therapy that deals with identifying, labeling, and exploring alleged "parts" of a an individual's personality and does not necessarily require a diagnosis of DID. Parts work is also used in such empirically unsupported practices as neurolinguistic programming (NLP).

In this chapter, we also review some of the changes and developments that have occurred in the continuing saga of New Age and other related novel, unsupported therapies. More constructively, we propose what might be done to place psychotherapies on a firmer scientific footing. Rather than attempt to catalogue all forms that these therapies take, which would require several volumes, we use energy psychology/Thought Field Therapy and some of the newer variants of RMT and DID therapy/parts work as exemplars, and discuss the key trends we have observed in clinical practice, education and continuing education of therapists, advertising and promotion, professional journals, and the legal arena.

NEW AGE AND NOVEL UNSUPPORTED THERAPIES

The term "New Age" is a vague, umbrella term that subsumes a number of nontraditional belief systems associated with unconventional mystical practices and therapies. Originally, the term referred to an astrological belief that society is undergoing a paradigm shift from the Piscean to the Aquarian Age (Ferguson, 1980). The New Age movement means many different things to different people. In the previous edition of this book, Singer and Nievod (2003) provided the following as examples, among others: beliefs in extraterrestrials abducting people and then repressing their memory of the event; belief in reincarnation; and the belief that trauma, including birth trauma, is the cause of all psychological problems and that one can regress people to trauma that occurred in infancy and past lives to eliminate psychological problems.

We would add that a belief that the body has an invisible energy system that contains meridians and chakras is an increasingly popular belief system held by many New Age adherents. Although the term "New Age" brings to mind stereotypes of eccentric people, this is not necessarily the case when it comes to modern proponents of such therapies, who may not hold the associated mystical New Age belief systems. Some advocates of New Age psychotherapies may engage in such practices for more pragmatic

reasons. One recent survey of mental health professionals found that the most common reason for using such therapies was apparent success with them in their clinical experience (Pignotti & Thyer, 2012). Of course, client improvement or anecdotes alone are a weak evidentiary reed. To determine if a therapy is effective, it is necessary to show that a given treatment, New Age or not, yields better results compared with no treatment or a credible placebo treatment. Many client conditions, physical or emotional, are self-limiting, and often clients report or experience improvement to reduce cognitive dissonance or to please the therapist, or in response to the strong but subtle social psychological pressures that they report benefits from treatment.

Not all proponents of energy therapies, RMT, or parts work identify themselves as New Agers. Accordingly, to avoid getting sidetracked into debates over semantics and labels that seem to inflame, rather than engage in a thoughtful discussion of important issues at hand, we (Pignotti & Thyer, 2009a) have coined a more neutral term: novel unsupported therapies (NUSTs; Pignotti & Thyer, 2012). Our intention in critically examining these approaches is not to attack or malign either the therapists using them or the clients receiving them. Indeed, one of us (Pignotti) was once an advanced practitioner of Thought Field Therapy and a close associate of its founder (see Pignotti, 2007, for a detailed account). The other author, Thyer, received counseling as a teenager from a New Age-style practitioner who connected him to a Mathison electropsychometer, developed by a chiropractor, Volney Mathison, who was a proponent of L. Ron Hubbard's Dianetics. The device was a precursor to Hubbard's E-meter, used in Scientology auditing sessions, a pseudoscientific form of psychotherapy (Mathison, 1954). In his teens Thyer was also active with a New Age organization called Concept-Therapy (Fleet, 1997; Wolff, 1968; *www.concept-therapy.org*). Although both of us left these groups many years ago, we believe that our earlier involvement in these approaches provides us with a useful perspective on how persuasive they can be to therapists and clients alike. Our intent is to provide readers with a better understanding of why mental health professionals are attracted to such interventions and how scientific thinking can improve treatment efficacy and minimize the risk of iatrogenic (harmful) treatment effects.

A Select History of New Age Thinking

The American Transcendentalist Ralph Waldo Emerson said, "Once you make a decision, the universe conspires to make it happen." This quote captures the essence of New Age thinking—the belief that by concentrating one's thoughts, one can cause changes in the physical world in accordance with one's affirmations and visualizations. This magical belief system has probably been around as long as humanity and became formalized in a movement called New Thought in the latter part of the 19th century.

Although the New Thought movement was an amalgam of religious and pseudoscientific beliefs, we focus on those aspects of New Thought concerned with how the mind can supposedly influence one's health and physical environment.

In his 1902 book, *The Varieties of Religious Experience,* American philosopher William James termed the movement "Mind-Cure," which held that one's ailments, physical and mental, were caused by false beliefs and could be counteracted by correct beliefs, affirmations, and visualization exercises. In 1906, William Atkinson published his influential book *Thought Vibration or the Law of Attraction in the Thought,* which contains such statements as:

- "We close our eyes to the mighty law that draws to us the things we desire or fear that makes or mars our lives."
- "When we think we send out vibrations of a fine ethereal substance, which are as real as the vibrations manifesting light. . . . That these vibrations are not evident to our five senses is no proof that they do not exist."
- "We are sending out thoughts of greater or less intensity all the time, and we are reaping the results of such thoughts. Not only do our thoughts influence ourselves and others, but they have a drawing power—they attract to us the thoughts of others, things, circumstances, 'luck,' in accord with the character of the thought uppermost in our minds." (cf. *http://gitacademy.tripod.com/GodsInTraining/ThoughtVibration.htm*)

New Thought and Mind Cure held that thinking, affirmations, and visualizations can actually create changes in the physical world. These movements further posited a pseudoscientific mechanism in which energy vibrations created in the brain produce these changes. For a simple example, if you wish to meet someone, concentrate on this desire and it will come true. Popular magazines such as *New Thought* and *Mind Cure* enjoyed wide circulation among the general public in the early 20th century. New Thought differs from traditional prayer in that prayer is usually directed to God/Jesus/Saints, requests are made for some form of intercession (e.g., healing, a change in the weather or wealth), and the "mechanism" is clearly religious or theologically based. In the New Thought and New Age traditions, in contrast, the mind/brain is said to generate vibrations or waves that can cause things to happen to oneself and in the outside world. This assertion possesses the patina of scientific legitimacy, despite the fact that the subtle type of brain vibrations postulated by New Thought has never been detected, and controlled experiments demonstrating that New Thought techniques actually improve one's health or changing the physical world appear to be lacking.

In 1920 the French psychologist Emile Coué published *Self-mastery*

through Conscious Autosuggestion, and Coueism became popular as a form of mantra recitation and therapy. At the beginning and end of every day, the client was to recite the phrase "Every day in every way, I am getting better and better." Coué claimed organic improvements in mood, attitude, and emotions, and he even stated that physical disease could be brought about though his methods. Later variations on this theme developed in the middle of the last century include Norman Vincent Peale's (1952) *The Power of Positive Thinking* and Napoleon Hill's (1937) *Think and Grow Rich.* More recently, the claim that thinking can create physical and mental health, and bring riches and other benefits was the centerpiece of the best-selling book *The Secret* (Byrne, 2006), which spawned a movie spinoff, additional books, and a television show. *The Secret*'s author, Rhonda Byrne, freely acknowledged that her inspiration came from the 1910 book *The Science of Getting Rich* by Wallace Wattles, itself a product of New Thought.

In the 1960s, the beliefs of New Thought and Mind Cure became amalgamated in the mystical hodgepodge called New Age. One contemporary practitioner noted that New Age workshops, seminars, and techniques had become big business: "The potential of the New Age movement .. . is a blurring of the lines between mystical and psychotherapeutic approaches, in which both systems become practiced in a more holistic way" (Fleischer, 2013, p. 3).

In the 1960s, parapsychologist Jose Silva developed the Silva Mind Control Method, a structured form of meditation involving affirmations and visualization that exemplifies many of the features of a New Age therapy (*www.silvamindbodyhealing.com/lpv*). The Silva mind control method became widely used and remains commercially available today. In the 1940s, Dr. Thurman Fleet of San Antonio, Texas, started a related movement, Concept-Therapy (see Fleet, 1997; Wolff, 1968). The continuation and commercial success of these New Age movements for over 50 years attest to their enduring appeal. Today, the Internet has enabled such therapies to be promoted and adopted by professionals more easily than ever. For example, the pseudoscientific treatment called the emotional freedom technique (EFT) links itself with the New Thought and New Age doctrine broadly called the "Law of Attraction." The Kindle book titled *Attract Abundance Now with EFT and the Law of Attraction* (Look, 2011) makes the following grandiose claims:

> By combining Law of Attraction exercises with the cutting-edge energy tool EFT (commonly referred to as "tapping") your energy must change—and you will quickly become a vibrational match to prosperity and success in all areas of your life .. . you will be able to raise your vibration dramatically and become an energetic match to exceptional success in all areas of your life—health, wealth and professional as well as personal relationships.

In this quote we see the invocation of "vibrations," changing one's personal vibration to attract wealth and health, and the melding of the Law of Attraction with the body tapping methods associated with some forms of energy psychology therapies.

For reasons that are unclear, psychotherapists specializing in trauma may be especially susceptible to the lure of New Age interventions. In a study of licensed clinical social workers (Pignotti & Thyer, 2012), we found that therapists specializing in trauma were three times more likely to practice NUSTs than were therapists with other specialties. Hence, we focus our discussion mainly on trauma therapies, with the caveat that some of these approaches are also sometimes marketed for a wide variety of psychological and physical problems. We begin with a discussion of how the practice of DID therapy and parts work has evolved over the past decade.

RMT, DID THERAPY, AND PARTS WORK

Some of the practices Singer and Nievod (2003) describe, particularly those practices used as part of RMT and DID treatment, have resulted in numerous lawsuits stemming from their potential iatrogenic effects. Several of these lawsuits have resulted in multimillion dollar judgments or settlements and loss or suspension of licensure for therapists (Grove & Barden, 1999). In an *amicus curiae* (friends of the court) brief, Harvard Psychology Professor Richard J. McNally noted, "The notion that traumatic events can be repressed and later recovered is the most pernicious bit of folklore ever to infect psychology and psychiatry. It has provided the theoretical basis for 'recovered memory therapy'—the worst catastrophe to befall the mental health field since the lobotomy era" (2006, p. 9).

In recent years, RMT therapies appear to have undergone a metamorphosis, resulting in a host of new untested alphabet therapies (therapies known by their acronyms) aimed at trauma survivors and practiced by a vocal minority of mental health practitioners who identify themselves as trauma specialists. In a recent handbook, clinical social worker Robin Shapiro (2010) described a panoply of such therapies that she termed an "alphabet soup" (p. 3) for which she provided a special acronym page following the table of contents. Not only did Shapiro present such untested therapies uncritically, but she also presented misinformation regarding therapies for trauma that boast the highest degree of support. For example, she claimed that exposure-based therapies for trauma make people worse, whereas the novel therapeutic approaches she described are superior. These assertions were based on clinical experience and anecdotes rather than research evidence. In the Introduction, Shapiro stated that "research is beyond the scope of most of this book" (p. 4); subsequent chapters bear this out, as most of the book is based on clinical anecdotes and descriptions rather than empirical data. Some of the NUSTs featured in Shapiro's volume include somatic

experiencing therapy (SE), sensorimotor psychotherapy (SP), somatic transformation, the specific trauma techniques of David Calof (a major and controversial figure in 1990s RMT), energy psychology, dyadic developmental psychotherapy (DDP), brainspotting and experiential integration, reenactment protocol, neurofeedback, and such parts work therapies as ego state therapy and internal family systems (IFS).

"Parts work" and controversial trauma therapy also figure prominently in a lawsuit. Between November 2001 and December 2012, four former patients of Castlewood Treatment Center, a residential treatment center for eating disorders, sued the center and its director for the induction of false memories involving trauma related to sexual and satanic abuse, and for the implantation of multiple personalities and diagnoses of DID (Nasseff, 2011; Taylor, 2012; Thompson, 2012; Travers, 2012). The four lawsuits contain similar allegations and resulted in confidential settlement agreements sealed by the court. The allegations that the therapist used hypnosis to create false memories of abuse, including the patients' belief that they were part of a satanic cult, as set forth in the complaints, bear similarities to the RMT cases of the 1990s. The therapist and the Castlewood Treatment Center have denied all charges.

An examination of Castlewood Treatment Center's website (*www.castlewoodtc.com/about/ifs-model*) reveals that it uses an eclectic array of trauma therapies, including IFS (Schwartz, 1995), a therapy that involves "parts" work that is used with clients with DID as well as clients with other diagnoses. IFS proposes that patients with a history of complex trauma (trauma that occurs over an extended period of time that involves multiple instances) develop subpersonalities of patients with DID and other conditions to cope with highly aversive events. IFS works with those proposed subpersonalities, the goal being to uncover and empower the real Self, which invariably is said to have only positive qualities. Any negative qualities are attributed to parts that need to be worked with. IFS, when successful, is said to result in an individual who is Self-led. Self, with a capital S, refers to a presumably completely healthy and wise part of a person's psyche. The goal of this therapy is to have this part be in charge of the individual's life and the other "parts."

Although IFS purports to be based on systems theory, the work of Sigmund Freud, and psychosynthesis (Assagioli, 1965), it fits what Singer and Nievod (2003) describe as New Age therapy goals of transformation and purification. The IFS "Self" is said to be untainted and healthy and to possess compassionate leadership qualities, once the subpersonalities or parts created by the person's abusive past are dealt with in therapy. It would seem that in IFS therapy, a simple statement by a client that she is feeling sad, angry, or fearful about something would be reframed by the therapist as the doings of a "part" that needs to be treated, rather than the client's true "Self."

In an apparently unpublished paper, posted on Castlewood's website

(Schwartz, Schwartz, & Galperin, n.d.), the authors admit, "Unfortunately, no well-constructed outcome studies testing the IFS model and methods have been completed," and they maintain that "the best evidence of IFS is from empirical observations in the clinician's office" (pp. 7–8). The authors assert that "until the results of these studies are in, skeptical clinicians are left to test these assertions within their own practices" (p. 8). This claim is concerning, given the vulnerability of the population this therapy targets (e.g., clients with eating disorders) and the potential for iatrogenic effects of this untested therapy, such as possible identity confusion and fragmentation of personality, as well as encouraging the recovery of potentially false memories, allegedly repressed by the "parts."

Nevertheless, this lack of evidence has not prevented IFS from marketing expensive training to mental health professionals. According to brochures on the main IFS website (see *www.selfleadership.org*), the training consists of three levels, which together cost $7,400. IFS also offers annual conferences (Internal Family Systems, 2011). One of the sessions led by a psychiatrist, entitled "Who's Taking What: The Integration of Psychopharmacology and Internal Family Systems," stated, "Therapists will learn how to work with parts in making medication decisions" (p. 9) and instructed participants on which medications work best for which "parts." It is troubling that medication is apparently being prescribed for "parts" in ways that have not been tested in randomized controlled trials.

RECOVERED MEMORY OF ABDUCTION BY ALIENS

In the version of this chapter in this book's first edition (Singer & Nievod, 2003), the authors discussed recovered memory of alien abductions. The authors presented a case study of a woman who with hypnosis came to believe that she had been abducted and sexually abused by aliens. Moreover, she concluded that this trauma was the cause of her presenting back pain problems. Over the past decade, a team of researchers (McNally et al., 2004) used script-driven imagery to examine the psychophysiological responses of people who had recovered memories of alien abduction. Selection of a highly improbable experience, namely, abduction by aliens, is an ethical and internally valid way to answer the question of whether people recover clearly false memories and experience the physiological responses and intense emotional responses associated with them.

Using newspaper ads, McNally and his colleagues recruited participants who believed they had recovered memories of being abducted by aliens ("abductees") and compared this group with a group of control participants who did not report such memories. Abductees prepared scripts of their memories and then listened to them, while their psychophysiological responses were recorded. Control group participants listened to tapes in which the abductees of the same sex and age described their experiences.

The two groups were also administered a battery of psychological tests. Participants in the abductees group scored significantly higher on measures of dissociation, absorption, and magical ideation than participants in the control group. Abductees also experienced greater physiological reactivity to the scripts compared with those in the control group, as well as heightened self-reported emotional responses. The abductees group displayed the kinds of physiological responses (e.g., increased heart rate and skin conductance) that people with verifiable trauma, such as war trauma, had displayed in other studies. This study counters the arguments made by RMT proponents that their memories must be true because they evoked vivid emotions and strong physiological responses, as it demonstrates that individuals with memories that are almost certainly false can also display such reactivity.

ENERGY PSYCHOLOGY AND THOUGHT FIELD THERAPY

Thought Field Therapy (TFT; Callahan & Trubo, 2001), a treatment developed by psychologist Roger Callahan, employs stimulation, usually by finger tapping, on purported acupressure points on the body. This procedure is performed while the client focuses on an emotionally disturbing issue, such as a traumatic event and/or its repercussions or fear. The early development of TFT began in the late 1970s when Callahan studied a chiropractic muscle testing technique called applied kinesiology, which he incorporated into TFT as an assessment technique. Callahan called this technique "causal diagnosis" and claimed that unlike conventional diagnoses used in mental health (e.g., standard DSM-5 diagnoses), his form of diagnosis addresses the root cause of all psychological problems, which he believes to be perturbations (i.e., energy disturbances). Callahan borrowed concepts from physics such as perturbation, field, and isomorphism, frequently citing physicist David Bohm (Bohm & Hiley, 1993). However, there is no credible evidence for such a human energy system or for perturbations. Because TFT displays the hallmark characteristics of pseudoscience, it is sometimes used as an exemplar of pseudoscientific practice (Herbert & Gaudiano, 2000).

As Singer and Nievod (2003) observed, the American Psychological Association had denied continuing education credits for TFT in the 1990s (Murray, 1999). They quoted Lilienfeld and Lohr's (2000, p. 5) optimistic statement that "these two recent actions herald a shift away from a laissez-faire approach to psychotherapeutic practice and toward heightened clinician accountability." This optimism appears to have been unwarranted.

In November 2012, the American Psychological Association reversed its position and approved the Association for Comprehensive Energy Psychology as a provider to offer training seminars in these therapies. Energy psychology is said to comprise a larger body of meridian tapping therapies

that include TFT and its many offshoots, the best known being emotional freedom techniques (EFT). One of the authors of this chapter (Thyer) served on the American Psychological Association's Continuing Education Committee when this latest request for approved provider status was slated to again come before the Continuing Education Committee for review and a determination was to be made. The Association for Comprehensive Energy Psychology contacted the American Psychological Association staff of the Continuing Education Committee to request that Thyer be recused from discussing their application, as he had written an article critical of the theory and practice of energy psychology (Pignotti & Thyer, 2009a). Thyer left the room for 20 minutes while the Association for Comprehensive Energy Psychology provider application was discussed. Upon his return, Thyer learned to his dismay that the Association for Comprehensive Energy Psychology had been approved as an American Psychological Association-endorsed provider of continuing education for psychologists. The Association for Comprehensive Energy Psychology is now providing training in energy psychology-based therapies, and licensed psychologists are acquiring skills in these methods, which produce effects essentially equivalent to placebo treatments (Pignotti & Thyer, 2009b).

An examination of the American Psychological Association's criteria for granting CE credits (American Psychological Association, 2009) provides some insight into how this decision may have been made. The American Psychological Association's (2012) *Standards and Criteria for Approval of Sponsors of Continuing Education* (American Psychological Association, 2012) describes four ways in which a CE provider can be approved by the American Psychological Association. One is by demonstrating that the methods taught are broadly accepted by the psychological practice, educational, or scientific communities. A second is that the content is supported by credible research; a third way is to demonstrate that the content is supported by peer-reviewed, noncommercial professional publications; and a fourth way is that the proposed content is related to professional ethical, legal, statutory, or regulatory issues impacting psychology.

Program content needs to meet only one of these standards, and because significant numbers of psychologists do practice some of these therapies, the first criterion was met. Also, some proponents of these approaches have created their own professional journals, which technically meets the third standard related to peer-reviewed publications.

The Association for Comprehensive Energy Psychology application is apparently not the first time the American Psychological Association has approved a provider of continuing education for psychologists who offer New Age therapies. The Kripalu Center for Yoga and Health, located in Stockbridge, Massachusetts, claims to be an American Psychological Association-approved provider of CE for psychologists (*http://kripalu.org/cecredits*). As we prepared this chapter, the Kripalu Center was advertising CE programming for psychologists on topics such as "Manage Your Mood

with LifeForce Yoga: I Am Bliss and So Are You (6 hours of CE)," "The Wise and Loving Heart: Meditation for Freedom and Compassion Everywhere (23.5 CE)," "Creating on Purpose: Manifesting through the Chakras (22 CE)," and "Seven Windows to the Soul: Transformation Through the Chakras (7 CE)." The last two programs are taught by a self-proclaimed pagan priestess. Here is how this workshop is described:

> This workshop will introduce you to the experience of the seven sacred energy centers, or chakras, that spin at the core of your being, generating your life patterns and physical health. Combining yoga and bioenergetic exercises with guided meditations, partner work, breath, sound, art, and dynamic discussions of chakra principles, you will gain valuable tools for self-diagnosis and healing of common energetic blocks.

Other continuing education programs approved for psychology continuing education at the Kripalu Center include "Breath–Body–Mind Level 1 Training for Transformation and Well-Being (10.5 CE)," "Energy Medicine: A Hands-on Experience (8.5 CE)," "Mind Whispering: A New Path to Freedom from Self-Defeating Emotional Habits (8.5 CE)," and "LifeForce Yoga and Internal Family Systems for Your Anxious Parts (8.5 CE)." This listing reflects the New Age integration of the secular (energy medicine) and the religious (yoga is derived from the Hindu religion. Professional psychologists can fulfill licensed-mandated CE requirements by taking such courses that are increasingly presented to the public as a legitimate form of psychotherapy.

Institutions of higher learning are also offering courses in TFT, EFT, and other meridian therapies. For example, the top-ranked School of Social Work at the University of Michigan, from which both authors of this chapter graduated, has offered a 2-day mini-course that included an uncritical presentation of EFT and other questionable therapies (Pignotti, 2007; see *www.ssw.umich.edu/shared/course_outlines/20065/bs790-002s06-1.pdf*). More recently, the reputable SUNY Buffalo School of Social Work announced on its Career Development webpage (see *www.socialwork.buffalo.edu/conted/trainings-buffalo.asp*) that in November 2012 it sponsored training course in another energy therapy, Reiki, entitled "Reiki I Certification: Using Energy Work in Human Service Practice." Reiki postulates the existence of a universal energy unknown to science and thus far undetectable surrounding the human body, which practitioners can learn to manipulate using their hands. Through use of both soft touch and touchless techniques, the goal is to restore a client's physical and mental health. Learning objectives for this continuing education program involves learning about Reiki energy, learning to experience and use this energy, and providing Reiki therapy to clients (School of Social Work, State University of New York, Buffalo, 2012, p. 3).

In 2012, another top-ranked school of social work program at SUNY

Albany accepted a $15,000 grant from the Association for Comprehensive Energy Psychology and the Global Gateway Foundation to conduct uncontrolled research on an EFT intervention, directed toward older adults who have developed PTSD following a heart attack. Of even greater concern is the fact that the website for the SUNY Albany School of Social Welfare includes unsupported claims about EFT: It maintains that EFT "calms the limbic structures of the brain" and is "grounded in neuroscience research" (see *www.albany.edu/ssw_enews/30924.php*).

Proponents of energy psychology are also making forays into peer-reviewed publications. Two favorable reviews were published in American Psychological Association journals, authored by psychologist David Feinstein (2008, 2012), a proponent of energy psychology. It is important to examine these reviews carefully because (1) unlike many of the studies, these reviews were published in high-impact American Psychological Association journals and (2) they purport to be accurate summaries of all the research on energy psychology (i.e., tapping therapies) to date and, as such, can be cited to imply that these practices are "evidence based" or "proven."

In his initial review, Feinstein (2008) claimed that energy psychology met the American Psychological Association's Division 12 criteria for a probably efficacious therapy for specific phobias and for weight loss. We (Pignotti & Thyer, 2009b) challenged a number of his claims, noting that his review omitted two randomized controlled studies (Pignotti, 2005; Waite & Holder, 2003) using sham control treatments that had produced null results and that Feinstein had only included reviews with favorable results. Also, he classified one study as a randomized controlled trial when it was only a clinical demonstration (Carbonell & Figley, 1999) that did not employ random assignment or formal tests of statistical significance. Because Feinstein did not clearly describe how he searched the literature, readers had no way of knowing the method he used to retrieve studies, including his inclusion and exclusion criteria. Moreover, we took issue with Feinstein's claims that energy psychology meets the American Psychological Association's criteria for an empirically supported treatment, due to flaws in the study methodology that we will discuss later in this chapter.

McCaslin (2009) also published a critique of Feinstein's (2008) review, noting a number of serious flaws in the research Feinstein had adduced in support of his claim that energy psychology is empirically supported. In one of the studies (Elder et al., 2007), there was no indication that dropouts were accounted for, nor were statistical outliers analyzed. The authors of the study reported that the treatment showed benefit over the control, but the results were not statistically significant. Also, McCaslin contacted the authors and found that participants had been allowed to come and go from the study as they pleased.

In his rejoinder to Pignotti and Thyer (2009) and McCaslin (2009), Feinstein (2009) asserted that the two controlled studies that had produced null results (Pignotti, 2005; Waite & Holder, 2005) actually provided

support for the therapies being tested. Waite and Holder's study included an EFT treatment group and three control groups: a group that received sham (not supposedly treatment-specific) acupressure points, a sham treatment group that tapped on an inanimate object (a doll), and a no treatment control group. The researchers found no statistically significant differences between the EFT treatment group and the two sham treatment control groups, although all three differed significantly from the no treatment group. Moreover, Feinstein attributed the null findings between treatment and sham controls to the possibility that acupressure points in the fingertips were stimulated during the sham treatment, when participants were asked to tap on a doll rather than their bodies. This explanation amounts to an ad hoc maneuver common to many pseudoscientific claims.

Feinstein also attempted to reverse the burden of proof, maintaining that Pignotti (2005) had not provided any evidence that nonspecific treatment effects could account for a 97% success rate. The study tested an advanced form of TFT known as voice technology (VT) by randomly assigning participants to a group that received actual VT or to a group that received a sham treatment (Pignotti, 2005). There were no significant differences between the "real" TFT treatment and the sham control group who received "fake" TFT. Hence, it is reasonable to conclude that nonspecific rather than specific treatment effects of VT were responsible for the statistically significant changes in participants' subjective distress. The burden of proof is on the claimant to show that TFT VT has specific treatment effects, not on the researcher to show that nonspecific treatment effects could have produced these changes. The fact that there were no differences between the VT and the sham treatment implies that something other than the treatment (e.g., nonspecific treatment effects, desire to please the therapist) could be responsible for the participants' positive self-reported changes or that another mechanism (e.g., exposure) could be responsible.

More recently, Feinstein (2012) published another review of energy psychology in an American Psychological Association journal in which he identified 18 randomized controlled trials and again claimed that energy psychology meets the American Psychological Association criteria for a research-supported treatment for posttraumatic stress disorder and other anxiety-based conditions. Nevertheless, the control group for 10 of the studies was a wait-list/no treatment condition, which does not rule out placebo effects or other nonspecific factors. Other studies used control conditions consisting of relaxation, diaphragmatic breathing, an inspirational talk by a coach, and supportive therapy, none of which directly control for the putative mechanism of action of the treatment, which involves the stimulation of specified acupressure points. Additionally, some of the studies were "in press" and unavailable for independent review.

The only study that compared two energy therapies with an active, empirically supported treatment (cognitive-behavioral therapy) for treating test anxiety included only five participants in each group. This limited the

ability to conduct an adequate statistical analysis. The only two studies included in the review that employed sham points or sequences (Pignotti, 2005; Waite & Holder, 2003) yielded null results, as previously discussed. Herbert and Gaudiano (2005) noted that these types of therapies are exceptions to the rule that double-blind studies cannot be conducted in psychotherapy and that a study could easily be designed, employing sham tapping points, in which both therapist and client are blind to the treatment condition. They further observed that "a trial using any lesser methodology than a single- or double-blind trial is largely uninformative" (p. 896). We agree that this is the case with the studies of lesser methodology adduced in support of TFT and EFT in Feinstein's review.

CONCLUSIONS

Feinstein (2012) provides some quotes illustrating the close link between New Age therapies and energy psychology, the injudicious amalgamation of legitimate science with mysticism:

> "energy healing practitioners believe they are also working with energies that involve a 'subtle' dimension that is not easily detected or measured" (p. 63). "Three subtle energy systems that may be addressed by energy psychology interventions: (a) energy pathways, such as 'meridians'; (b) energy centers such as 'chakras'; and (c) the energy field surrounding the body, known scientifically as the 'biofield' or in healing and spiritual traditions as the 'aura' " (p. 63). "Energy psychology interventions are believed to produce shifts in the energy systems that code psychological information, particularly the meridians, the chakras, and the biofield that surrounds the body" (p. 75).

Currently, conventional science has yet to validate the core principles of New Age psychotherapies—the idea that thoughts can influence one's external environment, the existence of subtle energies and fields—or of meridians, acupressure points, chakras, auras, or of the ability of some psychotherapists to reliably detect these constructs. The default assertion of New Age therapists is that because completely effective, inexpensive, and widely available treatments are not available for all client problems, they are justified in applying their unsubstantiated psychotherapies to clients. Here we invoke a principle of the Declaration of Helsinki of the World Medical Association (see *www.wma.net/en/30publications/10policies/b3*) as providing some possible ethical guidelines in this area. Specifically:

> 35. In the treatment of a patient, where proven interventions do not exist or have been ineffective, the physician, after seeking expert advice, with informed consent from the patient or a legally authorized representative, may use an unproven intervention if in the physician's judgment it

> offers hope of saving life, re-establishing health or alleviating suffering. Where possible, this intervention should be made the object of research, designed to evaluate its safety and efficacy. In all cases, new information should be recorded and, where appropriate, made publicly available.

We do not believe that the vast majority of the instances in which New Age or other NUSTs are applied by psychotherapists are consistent with the Helsinki Declaration's principle of providing informed consent, or of conducting research on the intervention's safety and efficacy. It is rare that the books, training workshops, CDs, or DVDs advertising training in these treatments, or offering them to the public as legitimate therapies, include a disclaimer along the lines of "The treatment being promoted lacks an adequate scientific evidence that it is an effective therapy. It is offered solely on the basis of the psychologist's clinical judgment, intuition, and personal beliefs."

Recall the confident assertion of one mental health professional who claimed, "I am a sensitive observer, and my conclusion is that a vast majority of my patients get better as opposed to worse after treatment." This professional was a psychiatrist who provided crude lobotomies on the brains of persons with mental illness during the 1950s (cf. Dawes, 1994, p. 48). It is now evident that prefrontal lobotomies are an ineffective treatment for persons with mental illness and in many instances are seriously injurious (Valenstein, 1986).

Here are some laudable ethical principles drawn from the field of medicine:

> The following general guidelines are offered to serve physicians when they are called upon to decide among treatments:
>
> (1) Treatments which have no medical indication and offer no possible benefit to the patient *should not be used*.
> (2) *Treatments which have been determined scientifically to be invalid should not be used*. (American Medical Association, 2012; emphasis in original. Retrieved March 27, 2013, from *www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics.page?*)

Offering a New Age or other NUST to a client, when the psychologist is aware that the proffered treatment lacks credible scientific evidence of its effectiveness, and when other psychosocial or medical interventions with a stronger evidentiary foundation exist, raises troubling ethical questions. We suggest that the ethical standards of the American Psychological Association and related professional organizations be amended to require full informed consent, especially when a practitioner wishes to provide an intervention lacking in empirical support, including New Age and many experimental treatments.

Dante claimed that false counselors were fated to reside in the eighth

circle of hell, with their heads on backwards, so they could only walk in reverse. We know of no empirical evidence to support Dante's conclusions, and we hope he is incorrect. However, his vision has as much support as do many of the claims made by the purveyors of New Age therapies. Perhaps some circumspection is in order.

GLOSSARY

Emotional freedom technique (EFT): An offshoot of TFT, developed by Gary Craig, a student of Roger Callahan, that is similar to TFT but differs in that the sequence of the treatment points is not believed to matter (*see also* Thought Field Therapy).

Energy psychology or energy therapies or meridian therapies: Interchangeable terms denoting any of a number of therapies that posit the existence of invisible energy fields surrounding the human body and an energy meridian system similar to that posited by proponents of acupuncture, which are said to influence a person's physical or mental health. Supposed misalignments in these fields are said to be correctable in various ways, such as tapping body points, touching, stimulation with weak electrical currents or laser pointers, or placing the therapist's hand on the meridian points.

New Age: An umbrella term used to describe nontraditional belief systems associated with unconventional mystical practices and therapies.

Novel unsupported therapies (NUSTs): Newly developed psychotherapies that either are under-researched or have been fairly well researched and shown not to be helpful, beyond placebo influences, yet make unsupported claims in their promotions for their efficacy and/or superiority over well-tested approaches.

Parts work: A term used to describe a number of different therapies (e.g., ego state therapy, internal family systems, transactional analysis) that are based on the belief that the human personality consists of multiple parts. This type of therapy is premised on the belief that it is therapeutic to address and deal with these parts as if they were separate entities.

Reiki: A form of energy therapy in which practitioners believe they can learn to manipulate the claimed human energy field by using their hands, with both soft touch and touchless techniques, as a means of restoring a client's physical and mental health.

Thought Field Therapy (TFT): A nontraditional form of therapy developed by psychologist Roger Callahan that employs finger tapping on specified points on the body in specific sequences while thinking the problem being addressed. It is claimed to rapidly address and cure a variety of psychological and physical problems.

REFERENCES

American Medical Association. (2012). *Code of medical ethics*. Retrieved March 27, 2013, from *www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics.page?*

American Psychiatric Association. (2009). *Opinions of the Ethics Committee on the principles of medical ethics*. Arlington, VA: Author. Retrieved March 29, 2013, from *www.psych.org/practice/ethics/resources-standards*.

American Psychological Association. (2009). *Standards and criteria for approval of sponsors of continuing education for psychologists*. Washington, DC: Author.

Assagioli, R. (1965). *Psychosynthesis: A manual of principles and techniques*. New York: Viking Press.

Atkinson, W. W. (1906). *Thought vibration or the law of attraction in thought*. Chicago: New Thought.

Bohm, D., & Hiley, B. J. (1993). *The undivided universe*. London: Routledge.

Braun, B. G. (1988). Continuum of dissociation. *Journal of Dissociation, 1*, 1.

Byrne, R. (2006). *The secret*. New York: Atria Books.

Callahan, R. J., & Trubo, R. (2001). *Tapping the healer within*. Chicago: Contemporary Books.

Carbonell, J. L., & Figley, C. (1999). A systematic clinical demonstration project of promising PTSD treatment approaches. *Traumatology, 5*, Article 4.

Coué, E. (1922). *Self mastery through conscious autosuggestion*. New York: American Library Service.

Dawes, R. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.

Elder, C., Ritenbaugh, C., Mist, S., Aickin, M., Schneider, J., Zwickey, H., et al. (2007). Randomized trial of two mind–body interventions for weight-loss maintenance. *Journal of Alternative and Complementary Medicine, 13*, 67–78.

Feinstein, D. (2008). Energy psychology: A review of the preliminary evidence. *Psychotherapy: Research, Practice, Training, 45*, 199–213.

Feinstein, D. (2009). Facts, paradigms, and anomalies in the acceptance of energy psychology: A rejoinder to McCaslin's (2009) and Pignotti and Thyer's (2009) comments on Feinstein (2008A). *Psychotherapy Theory, Research, Practice, Training, 46*, 262–269.

Feinstein, D. (2012, August 20). Acupoint stimulation in treating psychological disorders: Evidence of efficacy. *Review of General Psychology, 16*, 364–380.

Ferguson, M. (1980). *The aquarian conspiracy*. Los Angeles: J. P. Tarcher.

Fleet, G. T. (1997). *Where there is a will there is a way: A biography of Thurman Fleet, DC*. Boerne, TX: Concept-Therapy Institute. Available at *www.concept-therapy.org*.

Fleischer, P. (2013). Psychotherapy and mysticism in the New Age. Retrieved January 4, 2013, from *http://ecobites.com/eco-news-articles/natural-health/539-new-age-psychotherapies*.

Gaudiano, B. A., Brown, L. A., & Miller, I. W. (2012). Tapping their patient's problems away?: Characteristics of psychotherapists using energy meridian techniques. *Research on Social Work Practice, 22*(6), 647–655.

Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under *Daubert*/Kumho analysis. *Psychology, Public Policy and the Law, 5*, 234–242.

Herbert, J. D., & Gaudiano, B. A. (2005). Moving from empirically supported treatment lists to practice guidelines: The role of placebo concept. *Journal of Clinical Psychology, 61*, 893–908.

Hill, N. (1944). *Think and grow rich*. Meriden, CT: Ralston Society. (Original work published 1937)

Internal Family Systems. (2011). 2011 Internal Family Systems Conference. Retrieved December 29, 2012, from *www.selfleadership.org/downloads/IFS-BOSTON-2011-FINAL.pdf*.

James, W. (1985). *The varieties of religious experience*. Cambridge, MA: Harvard University Press. (Original work published 1902)

Lilienfeld, S. O., & Lohr, J. M. (2000, March/April). Thought field therapy practitioners and educators sanctioned. *Skeptical Inquirer,* p. 5.

Look, C. (2011). *Attract abundance now with EFT and the law of attraction*. Kindle book available at *www.amazon.com/dp/B00642W5JM*.

Mathison, V. (1954). *Creative imagery*. Los Angeles: Mathison Electropsychometer.

McCaslin, D. (2009). Comments and rejoinder: A review of efficacy claims in energy psychology. *Psychotherapy Theory, Research, Practice, Training, 46*, 249–256.

McNally, R. J. (2006). *Amicus curiae* brief in *Taus v. Loftus*, Supreme Court of California. Retrieved September 16, 2010, from *www.religioustolerance.org/rmtmcnally.htm*.

McNally, R. J., Lasko, N. B., Clancy, S. A., Macklin, M. L., Pitman R. K., & Orr, S. P. (2004). Psychophysiological responding during script-driven imagery in people reporting abduction by space aliens. *Psychological Science, 15*, 493–497.

Murray, D. J. (1999). APA no longer approves CE sponsorship for thought field therapy. *APA Monitor Online, 30*(11), 9.

Nasseff v. Castlewood Treatment Center, LLC and Mark Schwartz, Sc.D. (2011). Retrieved from *www.courts.mo.gov/casenet/base/welcome.do*.

Peale, N. V. (2003). *The power of positive thinking*. New York: Simon & Schuster. (Original work published 1952)

Pignotti, M. (2005). Thought field therapy voice technology vs. random meridian point sequences: A single blind controlled experiment. *Scientific Review of Mental Health Practice, 4*(1), 72–81.

Pignotti, M. (2007). Questionable interventions taught at top-ranked school of social work. *Scientific Review of Mental Health Practice, 5*, 78–82.

Pignotti, M., & Thyer, B. A. (2009a). The use of novel unsupported and empirically supported therapies by licensed clinical social workers. *Social Work Research, 33*, 5–17.

Pignotti, M., & Thyer, B. A. (2009b). Some comments on "Energy Psychology: A Review of the Evidence": Premature conclusions based on incomplete evidence? *Psychotherapy Theory, Research, Training, Practice, 46*, 257–261.

Pignotti, M., & Thyer, B. (2012). Novel and empirically supported therapies: Patterns of usage among licensed clinical social workers. *Behavioural and Cognitive Psychotherapy, 40*, 331–349.

Schwartz, R. (1995). *Internal family systems therapy*. New York: Guilford Press.

Schwartz, R., Schwartz, M., & Galperin, L. (n.d.). The internal family system model: Systematically-based individual psychotherapy for complex traumatic stress disorders. Retrieved December 23, 2012, from *www.castlewoodtc.com/wpcontent/uploads/2011/07/courtois_ctsd.pdf*.

Shapiro, R. (2010). *The trauma treatment handbook*. New York: Norton.

Singer, M. T., & Nievod, A. (2003). New Age therapies. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 176–204). New York: Guilford Press.

Taylor v. Castlewood Treatment Center, LLC. (2012). Retrieved from *www.courts.mo.gov/casenet/base/welcome.do*.

Thompson v. Castlewood Treatment Center, LLC and Mark Schwartz, Sc.D. (2012). Retrieved from *www.courts.mo.gov/casenet/base/welcome.do*.

Travers v. Castlewood Treatment Center, LLC. (2012). Retrieved from *www.courts.mo.gov/casenet/base/welcome.do*.

University at Albany School of Social Welfare. (2012). *2012–2013 Research guide*. Retrieved May 27, 2013, from *www.albany.edu.ssw*.

Valenstein, E. (1986). *Great and desperate cures: The rise and decline of psychosurgery and other radical treatments for mental illness*. New York: Basic Books.

Waite, W. L., & Holder, M. D. (2003). Assessment of the emotional freedom technique: An alternative treatment for fear. *Scientific Review of Mental Health Practice*, *2*(2), 20–26.

Wattles, W. D. (1910). *The science of getting rich*. New York: Wiley.

Wells, S., Polglase, K., Andrews, H. B., Carrington, P., & Baker, A. H. (2003). Evaluation of a meridian based intervention, emotional freedom techniques (EFT) for reducing specific phobias of small animals. *Journal of Clinical Psychology, 59,* 943–966.

Wolff, W. (1968). *Psychic self-improvement with concept-therapy: Complete story of an amazing new movement-healing with ideas*. New York: Grossett & Dunlap.

CHAPTER EIGHT

Constructing the Past

Problematic Memory Recovery Techniques in Psychotherapy

Steven Jay Lynn, Elisa Krackow,
Elizabeth F. Loftus, Timothy G. Locke,
and Scott O. Lilienfeld

In 1997, Nadean Cool won a $2.4 million malpractice settlement against her therapist in which she alleged that he used a variety of techniques and suggestive procedures to convince her that she had suffered horrific abuse and harbored more than 130 personalities, including demons, angels, children, and a duck (see also Lilienfeld & Lynn, Chapter 5, this volume). According to Ms. Cool, prior to therapy she had a few problems typical of many women, namely, a history of bulimia and minor depression. She initiated therapy to deal with her feelings concerning the sexual assault of a family member.

During 5 years of treatment in which her therapist allegedly insisted she could not improve unless she excavated traumatic past experiences that she had repressed, Nadean "recovered" what her therapist claimed were repressed memories of her having been in a satanic cult, of eating babies, of being raped, of having sex with animals, and of being forced to watch the murder of her 8-year-old friend. These memories surfaced after Nadean participated in repeated hypnotic age regression and guided imagery sessions and was subjected to an exorcism and 15-hour marathon therapy sessions. Her therapist charted her various supposed personalities and their dynamic interactions with one another, consistent with the symptoms of dissociative identity disorder (DID; formerly called multiple personality

disorder), which she seemingly developed in response to suggestive memory recovery techniques.

According to Ms. Cool, as her therapy progressed and she became overwhelmed by frightening images of events she came to believe occurred in her past, her psychological equilibrium deteriorated apace. Eventually, she began to have serious concerns about what had transpired in therapy, came to doubt that the memories she recovered were "real," terminated treatment with her therapist, and recouped much of the ground she had lost over the past 5 years. Dramatic legal cases such as Nadean Cool's provide vivid and cautionary examples of how highly directive and coercive therapeutic approaches can shape a patient's history and sense of personal identity.

Cases like Nadean's prompt two major questions. First, does the broader community of therapists use special techniques to enhance recall of traumatic events, such as child abuse? Second, do these techniques do more than merely access memories but, instead, shape them and thereby instate a distorted personal history? With respect to the first question, there is good reason to believe that the use of memory recovery techniques is not limited to rare yet highly publicized legal cases. Poole, Lindsay, Memon, and Bull (1995) surveyed 145 licensed U.S. doctoral-level psychotherapists who were randomly sampled from the National Register of Health Service Providers in Psychology in two studies and 57 British psychologists who were sampled from the Register of Chartered Clinical Psychologists. They found that 25% of the respondents who conduct therapy with adult female patients believe that memory recovery is an important part of treatment, believe that they can identify patients with repressed or otherwise unavailable memories as early as the first session, and use two or more techniques such as hypnosis and guided imagery to facilitate the unearthing of repressed memories. Additionally, Poole and colleagues reported that over three-quarters of the U.S. doctoral-level psychotherapists sampled in their study reported using at least one memory recovery technique to "help clients remember childhood sexual abuse." Polusny and Follette (1996) conducted a survey of 1,000 psychologists who were randomly selected from Division 12 (Clinical) and Division 17 (Counseling) of the American Psychological Association. Of the sample of 223 individuals who ultimately completed the survey, 25% of therapists reported using guided imagery, dream interpretation, bibliotherapy regarding sexual abuse, and free association of childhood memories as memory retrieval techniques from clients who had no specific memories of childhood sexual abuse.

In a study of 220 Canadian practicing mental health professionals, including 76 psychologists, Legault and Laurence (2007) found that 41% of psychologists agreed that "hypnosis enables people to accurately remember things they otherwise would not" (p. 121) and that a remarkable 67% of psychologists agreed that "hypnosis can be used to recover memories of actual events from as far back as birth" (p. 121). Moreover,

27% of psychologists endorsed the view that "recovered memories must be reliable because no wants to have been abused as a child" (p. 122). The authors reported that 22% of participants endorsed the use of hypnosis for memory recovery and 20% endorsed age regression for this purpose. More recently, Pignotti and Thyer (2011) reported that of 368 U.S. social workers whom they surveyed (1) 7.6% reported using age regression for the treatment of sexual abuse, (2) 2.5% reported using past lives therapy, and (3) 9.8% reported using traumatic incident reduction, a technique that involves experiencing purportedly repressed memories in a safe and comfortable environment. None of these interventions enjoys any research support.

Given the apparent ubiquity of memory recovery techniques and the problematic rationale that has been advanced for their use, it is imperative to address the second question of whether memory recovery techniques that are used in therapy to recover childhood memories of abuse can distort clients' recollections of past events and create memories of experiences that never occurred. In this chapter, we review research on the impact of suggestion on memory, as well as the effects of problematic therapeutic procedures of symptom interpretation, hypnosis, dream interpretation, and bibliotherapy on memory.

SUGGESTING FALSE MEMORIES

If memories of highly traumatic events were truly repressed, there would be some justification for the use of special techniques to promote their recovery. However, there is no compelling foundation for the claim that memories are repressed in the face of trauma (McHugh, 2008; McNally, 2005). To the contrary, traumatic events are generally more memorable than memories of mundane events, but remain subject to reconstructive processes just like nontraumatic memories. Moreover, when memories are "recovered"—that is, people report having recovered memories—it may simply mean that they had not thought about the events for some time, or had forgotten their earlier recollection and remembered the event spontaneously upon encountering a reminder inside or outside of psychotherapy (McNally & Geraerts, 2009).

Given serious concerns about the theory of "repressed memories," recent years have witnessed a strong backlash against the use of memory recovery techniques in psychotherapy, and much of this pushback has centered on the use of suggestive methods in treatment (e.g., hypnosis). Still, disturbingly high rates of clinicians continue to endorse use of the problematic memory recovery techniques we review in this chapter. In the next section, we present research that supports the idea that it is not difficult to implant memories of a variety of childhood experiences. In later sections, we review more specialized procedures for enhancing memories that are, not infrequently, used in the context of psychotherapy.

We begin our discussion with the seminal research of Elizabeth Loftus and her colleagues and research that converges on the conclusion that people can be led to integrate an entirely fabricated event into their personal histories (Loftus, 1993; Loftus & Ketcham, 1994; Loftus & Pickrell, 1995). In Loftus's research, participants were asked to remember real and fictitious events (e.g., getting lost in a shopping mall). A parent or older sibling initially provided a few details to help create the false event, such as where the event occurred. All subjects participated in interviews over several days. Subjects claimed to remember the false event and provided surprisingly detailed accounts of the event that they believed actually occurred.

Similar studies in other laboratories also found that a significant minority of people reported false events. For instance, Hyman, Husband, and Billings (1995) asked college students to recall childhood experiences that had been recounted by their parents. In addition to actual events reported by parents, each participant was provided with a false event—either an overnight hospitalization for a high fever and a possible ear infection, or a birthday party with pizza and a clown—that purportedly occurred at about the age of 5. The parents confirmed that neither of these events actually took place. Although none of the participants recalled the false event during the first interview, 20% said they remembered some details of the false event in the second interview.

Along with true events, Hyman and colleagues (1995) presented different false events including accidentally spilling a bowl of punch on the parents of the bride at a wedding reception and having to evacuate a grocery store when the overhead sprinkler systems erroneously activated (Hyman et al., 1995). Again, none of the participants recalled the false event during the first interview, but 18% remembered something about it in the second interview and 26% in the third interview.

In a second study by Hyman and Pentland (1996), participants who engaged in guided imagery reported more false memories of attending a wedding and knocking over a punchbowl than did individuals in a control group who received instructions to do their best to remember childhood events. By the third recall trial, 9% of the control group reported the false event compared with 25% of the guided imagery group.

Porter, Yuille, and Lehman (1999) found that 26% of participants reported at least one "complete" (i.e., incorporated all misinformation into memory) false memory of six suggested emotional childhood events (serious animal attack, serious indoor accident, serious outdoor accident, getting lost, serious medical procedure, being injured by another child). The events were suggested in three different interviews over a 2-week period in which the interviewer attempted to elicit false memories by using guided imagery, context reinstatement, and mild social pressure, and by encouraging repeated recall attempts. When partial memories were considered in which some of the information was recalled or the individual was uncertain

about whether the memory was false, more than half of individuals who were exposed to a variety of memory recovery techniques could be led to report an emotional false memory, with a substantial minority elaborating the memory in a detailed and confident manner.

Research also supports the idea that people can genuinely believe in memories of traumatic events that are highly unlikely to have occurred in real life such as being abducted by aliens (McNally, 2012; Spanos, Burgess, & Burgess, 1994). Otgaar, Scorboria, and Smeets (2012) were able to instate memories of being abducted by an unidentified flying object (UFO) when they increased the plausibility of the event by providing children aged 7–12 with newspaper articles that suggested such experiences were not uncommon. McNally et al. (2004) found that individuals who believed they had experienced a UFO abduction experienced pronounced elevations in psychophysiological arousal (i.e., heart rate, skin conductance, and frontalis EMG responses) symptomatic of posttraumatic stress disorder when they were exposed to script-driven imagery of their reported UFO abductions. This finding suggests that improbable "memories" can be accompanied by intense emotions that enhance their realistic quality.

Cultural narratives probably play a role in claims of satanic ritual abuse—for which there is no evidence the claims are veridical—and symptoms of DID. In a demonstration of the role of cultural beliefs, Stafford and Lynn (2002) asked participants to role-play symptoms of either DID, major depression, or adjustment problems. The researchers found that the DID role players were more likely to report they experienced satanic ritual abuse during childhood, compared with individuals who role-played the other psychological symptoms. The researchers contended that cultural scripts of DID as a posttraumatic condition could account for the reports of satanic ritual abuse among DID role players and the virtual absence of such reports among other role players.

False Feedback: Bogus Personality and Early Experiences

For ethical reasons, researchers have not directly tested the hypothesis that false memories of history of abuse can be elicited by informing individuals that their personality characteristics are suggestive of such a history. However, studies have shown that it is possible to instate implausible or false memories of other events by providing individuals with bogus information about their personality characteristics. Spanos, Burgess, Burgess, Samuels, and Blois (1999) were able to convince participants that their test scores revealed that they were "High Perceptual Cognitive Monitors."After participants were informed falsely that the study was designed to recover memories in order to confirm the personality test scores, the experimenter age regressed individuals back to the crib following either hypnosis or nonhypnotic age regression instructions. In the nonhypnotic group, 95% of the participants reported infant memories and 56% reported the target mobile.

However, all of these participants indicated that they thought the memories were fantasy constructions or they were unsure if the memories were real. In the hypnotic group, 79% of the participants reported infant memories and 46% reported the target mobile. In addition, 49% of these participants believed the memories were real and only 16% classified the memories as fantasies.

DuBreuil, Garry, and Loftus (1998) administered a test to college students that purportedly measured personality and were told that, based on their scores, they were likely to have participated in a nationwide program designed to enhance the development of personality and cognitive abilities by means of the use of red and green moving mobiles. Some participants were told that this enrichment occurred in the hospital immediately after their birth, whereas other participants were told that the mobiles were placed in kindergarten classrooms. Finally, participants were given the false information that memory functions "like a videotape recorder" and that memory retrieval techniques (e.g., nonhypnotic age regression) can access otherwise inaccessible memories, after which they were age regressed to the appropriate time period (i.e., kindergarten or infancy) with suggestions to visualize themselves at the target age. Twenty-five percent of the kindergarten group and 55% of the infancy group reported the target memory. All kindergarten participants believed their memories corresponded to real events. In the infancy group, 33% believed in the reality of their memories, 50% were unsure, and 17% did not believe in the reality of their memories. Finally, a pretest questionnaire revealed that participants who believed that specific techniques could recover memories reported more suggested memories.

More recently, Berkowitz, Laney, Morris, Garry, and Loftus (2012) provided participants with a faked computer profile presented as unique personalized information. To establish a positive or negative emotional tone, the profile supposedly confirmed their experience of either pleasant childhood activities (e.g., watching cartoons, playing with friends in the positive condition) or common childhood fears (e.g., loud noises in the negative condition). Next they received newspaper excerpts of stories of supposed personal relevance to them regarding a trip to Disneyland in childhood associated with positive experiences with the cartoon character Pluto (e.g., Pluto delighted children by licking their ears in the positive condition) or negative experiences with Pluto (e.g., Pluto licked the children's ear in an unpleasant manner; Pluto abused hallucinogenic drugs in the negative condition). Compared with participants who received no specific information about Pluto, participants in the experimental conditions expressed more confidence that Pluto had licked them. Moreover, in the positive condition, participants were more likely to pay more money for a Pluto toy compared with their willingness to pay for the toy the previous week, reflecting the influence of the implanted memory on current adult behavior. In another study using false feedback based on "personality profiles," 17.9%

of participants reported false memories of punching a person and giving the person a black eye (Laney & Takarangi, 2013). These studies suggest that providing individuals with information about their personalities, which often occurs in psychotherapy, may influence not only their reports of childhood memories but also their behaviors in adulthood.

Researchers have used questionnaire-based feedback to participants to manipulate memories of childhood preferences for food (e.g., Bernstein & Loftus, 2009; Geraerts et al., 2008). Studies have shown that childhood memories implanted in adults in the lab related to food preferences (e.g., getting sick eating egg salad—Geraerts et al., 2008; getting sick eating strawberry ice cream—Bernstein, Laney, Morris, & Loftus, 2005) can influence current behaviors related to such preferences (e.g., subjects ate fewer egg salad sandwiches in the lab) for as long as 4 months after the false suggestion (Geraerts et al., 2008). Interestingly, these memories can be virtually indistinguishable from true memories of childhood experiences (see Morris, Laney, Bernstein, & Loftus, 2006).

Dream Interpretation: Expert Personalized Suggestion

With Freud (1900/1953, 1918/1955) as an ardent proponent, dream interpretation became a staple in the panoply of psychoanalytic techniques. Viewed as the "royal road to the unconscious," dreams have been used to provide a window on past experiences, including supposed repressed traumatic events. For example, van der Kolk, Britz, Burr, Sherry, and Hartmann (1994) claimed that dreams can represent "exact replicas" of traumatic experiences (p. 188), a view not unlike that propounded by Fredrickson (1992), who argued that dreams are a vehicle by which "buried memories of abuse intrude into . . . consciousness" (p. 44).

A survey based on the responses of members of the American Psychological Association (Fox, 2002) indicates that 90% of respondents rated dream interpretation to be at least slightly effective in treatment and 83% used dreams to some extent. These statistics are of particular interest given that no data exist to support the idea that dreams accurately reveal autobiographical memories that fall outside the purview of consciousness (Lindsay & Read, 1994). When dreams are interpreted by clinicians as indicative of a history of child sexual abuse (Bass & Davis, 1988; Fredrickson, 1992), the fact that the information is provided to the client by an authority figure can constitute a strong suggestion that abuse, in fact, occurred in "real life."

Ethical constraints preclude studies that examine false memories in the context of child abuse. However, Mazzoni and her colleagues (Mazzoni, Lombardo, Malvagia, & Loftus, 1999b) conducted a series of studies that attempted to simulate the effects of dream interpretation of nonabuse-related yet stressful life events. In their first study (Mazzoni et al., 1999b), participants reported on their childhood experiences on two occasions,

separated by 3–4 weeks. Between these sessions, an expert clinician analyzed a dream report that participants had brought to the session. No matter what the content of their dreams, the participants received the suggestion that their dream was indicative of having experienced, before age 3, such events as having been lost in a public place or abandoned by their parents. Although participants had previously indicated that they had not experienced these critical events before age 3, this 30-minute therapy simulation led many individuals to develop new beliefs about their past. Relative to controls who had not received the personalized suggestion, these "therapy" participants were far more likely to develop false beliefs that before age 3 they had been lost in a public place, had felt lonely and lost in an unfamiliar place, and had been abandoned by their parents. In another study using a similar procedure, Loftus and Mazzoni (1998) showed that participants whose dreams were interpreted to indicate that they had experienced a very dangerous event before age 3 later reported an increased belief that the dangerous event had occurred. The new beliefs that were generated by the dream interpretation were maintained for at least 4 weeks

In a third study, Mazzoni, Loftus, Seitz, and Lynn (1999a) provided participants in the experimental condition with specific feedback suggesting that their dreams indicated they had had an interaction with a bully or had been lost in a public place. Control participants were either given a brief lecture about dreams by the "clinical psychologist" or did not attend the middle session. The authors found that the dream interpretation increased participants' confidence that the target event had occurred compared with control participants. Six of the 22 (27%) participants in the dream interpretation condition recalled the bullying event and 4 out of the 5 (80%) participants in the dream interpretation condition recalled the getting-lost event. Overall, 10 out of 27 participants (37%) in the experimental condition reported the target memory, although the ages given for the target event tended to be over 3. In conclusion, a growing body of evidence demonstrates that it is possible to implant autobiographical childhood memories using a variety of strategies that involve personality and dream interpretation.

Physical Symptom Interpretation: Body Memories

A "body memory" can be defined as an unexplained physical symptom that is interpreted as the result of childhood sexual abuse (CSA) or other historic trauma (cf. Frederickson, 1992; Levis, 1995). In van der Kolk's (1994) words, "The body keeps the score" (p. 253). According to survey research, 36% of U.S. psychotherapists two decades ago (Poole et al., 1995) interpreted body pains or physical symptoms as indicative of a history of childhood sexual abuse.

Some surveys of women seeking medical services have found that abuse survivors do indeed report more frequent physical symptoms, visits

to physicians, and lifetime surgeries than do women without a history of abuse (Austin, 1995). Austin (1995) attempted to explore the correlations of location of physical symptoms in adult survivors of CSA with the location of the purported physical insults during the abuse. Although a significant correlation was found for gynecological symptoms in adult sexual abuse survivors (i.e., adult rape), this association was not evident for CSA survivors. Relatedly, there were no significant correlations for gastrointestinal, throat, or rectal bleeding symptoms in adult or childhood survivors of sexual abuse. Nevertheless, these findings were limited by a small sample size ($N = 14$). Hence, although the literature suggests a relative abundance of physical symptoms in CSA survivors, the limited available research does not reveal a correlation between location of abuse and physical symptoms, thereby providing no clear support for the "body memory" hypothesis.

At least one treatment approach is based on the notion of body memories. Levis (1995) instructed patients to focus on a particular physical sensation or pain and to bridge the experience back in time to the etiology of the pain. This approach clearly implies that current physical problems have a historical psychological or physical etiology and engenders clear expectancies regarding the antecedents of current problems based on symptom interpretation. It is notable that similar expectancy-altering information has elicited highly implausible memory reports (e.g., abuse in a "past life"; Spanos, 1996) in previous research. Studies that use the concept of "body memories" to suggest that individuals experienced specific early life events are warranted to examine the effects of this interpretive set on the creation of false memories.

SEARCHING FOR MEMORIES: MEMORY RECOVERY METHODS

The term "memory work" has been used to refer to psychotherapy that focuses on retrieving "repressed" memories of childhood sexual abuse (Loftus, 1993). The memory recovery techniques we review are typically used with two kinds of clients: those who remember abuse but wish to fill in more details, and those who suspect early childhood abuse but have no memories of maltreatment.

Uncovering Early Memories: Direct and Repeated Questioning

The most basic technique that clinicians use to promote recall of early childhood events is to ask direct questions to prod memory. Many clinicians have long regarded the exploration of very early memories as crucial to the enterprise of psychotherapy (Bindler & Smokler, 1980; Papanek, 1979). Olson (1979) articulated a belief shared by many therapists (Papanek, 1979) that "[early memories] when correctly interpreted often reveal very quickly the basic core of one's personality, or life-style, and suggest

important, bedrock themes with which the therapist must currently deal in treating the client" (p. xvii).

Although certain early memories might well have special significance, such memories are highly malleable. By examining early memory reports, it is possible to study the influence of memory recovery techniques on implausible memories that cross the threshold of infantile amnesia, which is about 2 years of age. The inability to recall very early life events in adulthood can be attributed to development of language and the self, and structural changes in the brain necessary to support the acquisition and storage of complex events (Nelson & Fivush, 2004; Howe, 2003). Most adults' earliest reported memories date back to between 36 and 60 months of age; virtually all contemporary memory researchers agree that accurate memory reports of events that occur before 24 months of age are extremely rare (see Malinoski, Lynn, & Sivec, 1998, for a review). Even childhood memories of stressful events (e.g., emergency room visits) assessed 5 years later show different outcomes as a function of age (Peterson, 2012): Children approximately age 26 months tend to show long-term recall of the event 5 years later; below 26 months, 17 of 27 children did not report any recall of the event. The recall of the remaining 10 children contained more errors than accurate remembrances. Spanos (1996) argued that "the phenomenon of infantile amnesia is relevant to the topic of recovered memories in that it suggests that memories recovered in therapy, of abuse that supposedly occurred when the person was younger than 3 or so years of age, are very likely to be confabulations" (p. 80), misdated recollections, or stories heard about early events (Bruhn, 1984; Loftus, 1993).

Very early memory reports, like reports of later childhood events, are vulnerable to subtle suggestive influences. Lynn, Malinoski, and Green (1999) used two different wordings to try to elicit early memories. In one version, the high expectancy case, the participants were told, "Tell me when you got an earlier memory." In the other version, the low expectancy case, participants were told, "If you don't remember, it's all right." The high expectancy version led to earlier memory reports: The high expectancy mean was 2.48 years, whereas the lower expectancy mean was 3.45 years, a difference of nearly 1 full year. By the end of four recall trials, 43% of participants in the high expectancy condition reported a memory at or before 2 years of age, compared with 20% of participants in the low expectancy group.

If subtle suggestions that convey the expectation that earlier memories can be remembered result in marked changes in the age of earliest reported memory, what would be the effect of using memory recovery techniques to elicit early memories? Malinoski and Lynn (1999) addressed this question in a study in which interviewers repeatedly probed for increasingly early memories until participants twice denied any earlier memories. Participants then received a strong suggestion to promote earlier memory reports using "memory recovery techniques" similar to those promoted by some

therapists (e.g., Farmer, 1989; Meiselman, 1990). Interviewers asked participants to close their eyes, see themselves "in their mind's eye" as a toddler or infant, and "get in touch" with memories of long ago. Interviewers also conveyed the expectation that it was possible to recall very early life events by informing participants that most young adults can retrieve memories of very early events—including their second birthday—if they "let themselves go" and try hard to visualize, focus, and concentrate. Interviewers then asked for subjects' memories of their second birthdays. Participants were complimented and otherwise reinforced for reporting increasingly early memories.

The mean age of the initial reported memory was 3.70 years. Only 11% of individuals reported initial memories at or before age 24 months, and 3% of the sample reported an initial memory from age 12 months or younger. However, after receiving the visualization instructions, 59% of the participants reported a memory of their second birthday.

After the birthday memory was solicited, interviewers pressed participants for even earlier memories. The mean age of the earliest memory reported was 1.60 years, fully 2 years earlier than their initial memory report. One of the most interesting findings was that 78.2% of the sample reported at least one memory that occurred at 24 months of age or earlier. Furthermore, more than half (56%) of the participants reported a memory between birth and 18 months of life, a third of the participants (33%) reported a memory that occurred at age 12 months or earlier, and 18% reported at least one memory of an event that occurred at 6 months or earlier, well outside the boundary of infantile amnesia. Finally, a remarkable 4% of the sample reported memories from the first week of life. Clinical practitioners should also be alert to the fact that repeated questioning/retrieval attempts can enhance suggestibility (Chan & LaPaglia, 2011) and lead individuals to perceive that they have gaps in their memory. This practice can also foster the impression that their memory gaps may be associated with the failure to remember repressed traumatic events (see Belli, Winkielman, Read, Schwarz, & Lynn, 1998).

Guided Imagery and Imagination

Clinicians often use special techniques to promote recall of early childhood events that go well beyond repeated questioning. Lindsay and Read (1994) described guided imagery as a technique whereby patients are instructed to relax, close their eyes, and imagine various scenarios described by the therapist. However, concerns have been expressed (Lindsay & Read, 1994; Loftus, 1993) about such highly controversial approaches as the use of imagery procedures in therapeutic settings to elicit recall of allegedly repressed or dissociated memories of childhood sexual abuse. For example, Roland (1993) proposed using a visualization technique for jogging

"blocked" memories of sexual abuse and a "reconstruction" technique for recovering repressed memories of abuse.

Yet numerous studies now show that asking people to imagine events can create false memories or increase confidence in the likelihood that particular childhood events occurred. This latter phenomenon is known as imagination inflation (Garry, Manning, Loftus, & Sherman, 1996; Garry, Sharman, Wade, Hunt, & Smith, 2001; Heaps & Nash, 1999; Sharman & Scoboria, 2009). For example, Mazzoni and Memon (2003) asked participants to imagine having experienced a medical procedure (i.e., "Having a nurse remove a skin sample from my little finger") that was not performed in the United Kingdom when participants were children, a methodology that guarantees memory reports of the imagined event were false. Control participants did not imagine the medical procedure, but read a passage about the false event and responded to questions about it. Participants who imagined the event expressed more confidence that the event actually occurred when they were retested a week later. Further, when asked to provide memory reports of the event, 40% of participants who imagined the event provided false memory reports, compared with 23% of participants who read passages. Other studies show that imagining events, even implausible events (e.g., proposing marriage to a cola machine) on a single occasion for a brief period of time (10 seconds), can create false memories (Seamon, Philbin, & Harrison, 2006).

Lampinen, Odegard, and Bullington (2003) found that some participants reported incorrectly that they had engaged in actions they only imagined. Moreover, over succeeding imagination trials, their false recollections more closely approximated true recollections in terms of sensory details, emotion, contextual details. When a nonperformed action was imagined five times, differences between true and false memories were smaller, yet still evident.

Krackow and Rabenhorst (2010) demonstrated that when participants were asked to narrate imagined events they previously denied, their level of physiological reactivity increased from baseline to a level comparable to that observed in participants who narrated childhood events they earlier claimed they experienced. The researchers hypothesized that physiological reactivity during imagination is one contributor to the development of false memories. Using this same dataset, Krackow (2010) found that narratives of the imagined events could be distinguished from narratives of the experienced events: Imagined events contained a higher percentage of emotion words, whereas experienced events were lengthier and contained a higher percentage of cognitive processes words than imagined events. If emotion enhances the credibility of false memories, then therapists or onlookers may mistakenly conclude that false memories are, in fact, veridical.

Taken together, the findings reviewed show that imagination can create false memories of even implausible events after a single brief episode. The

available evidence raises caution regarding interventions in psychotherapy that use fantasy, imagination, and guided imagery to promote recall.

Using Photographs to Cue Memory

Memory recovery therapists often use family photographs to cue childhood images and "repressed" memories of suspected abuse (Lindsay, Hagen, Read, Wade, & Garry, 2004). Accordingly, researchers have examined the question of whether doctored photographs can create false memories. Wade, Garry, Read, and Lindsay (2002) interviewed adults on three occasions during which they were asked about a combination of true and false childhood events. One particular false event was of interest, a hot air balloon ride in childhood, depicted in a doctored photograph. If participants did not produce a memory for the hot air balloon ride, context reinstatement and guided imagery were employed as memory enhancement techniques. By the end of the third interview, 50% of participants reported complete or partial false memories. Moreover, participants with more elaborate false memories were more confident in the false event compared with the true events.

In a similar study, Lindsay and colleagues (Lindsay et al., 2004) asked parents to provide two true event narratives along with one school classroom photo for each year the reported two true events occurred. The researchers used mental context reinstatement and guided imagery to facilitate participants' recall of the events. All participants were also asked about the same researcher-devised false event (putting slime in the teacher's desk). Participants were asked to focus on recalling the false event in the next few days and to use the photograph as a memory cue. One week later participants returned to the lab and provided their memory reports. Results showed that 65.2% of participants in the photo condition reported memories of the false event compared with 45.5% of participants in the no-photo condition. The high percentage of participants in the no-photo condition that recalled false memories can be attributed to the combination of context reinstatement and guided imagery, as well as the expectation that participants would be able to recall the fabricated event. Taken together, these studies imply that photographs can cue false memories of childhood events.

In the above studies, imagined false memories were acquired over time, and additional memory enhancement procedures beyond photographs were employed. Strange, Garry, Bernstein, and Lindsay (2011) attempted to ascertain whether photographs alone could engender false memories immediately after participants viewed them. Participants either read newspaper headlines of true world events that actually occurred or read headlines of fabricated events, accompanied either with photographs only tangentially related to the headlines or no photographs. The combination of photographs and fabricated headlines led people to immediately report memories of the false news events.

In another study, Wade, Garry, Nash, and Harper (2010) varied whether an event narrative of a false childhood event (i.e., a hot air balloon ride) was provided before or after a doctored photo of the balloon ride. People exposed to the narrative of the event first were more likely to report false memories and images compared with people who saw the photograph first. Accordingly, information presented first was most influential.

Gerrie, Belcher, and Garry (2006) examined whether doctored photographs can engender recall of false details of true events. The researchers found that participants falsely recalled schema-consistent details when presented with video clips after viewing a movie. That is, they filled in the blanks with actions that typically happen and were more susceptible to memory errors on noncritical items. In summary, the aforementioned studies imply that clinicians should not use photographs as a memory recovery or enhancement technique. Moreover, special caution should be exercised in combining photographs with other memory enhancement techniques.

Event Discussion

Clinical practitioners would do well to heed research showing that memory has a social aspect: In some real-life cases, memories of sexual abuse occur after discussion with siblings (Garry, French, Kinzett, & Mori, 2008), and psychotherapy itself pivots on ongoing discussion with the therapist and interpretation of significant events. Several studies have examined how discussion influences true and false memories. In one study, French and colleagues (French, Sutherland, & Garry, 2006) implanted memories of false events individually in siblings through the use of photographs and narrative descriptions. Afterward, researchers asked the siblings to engage in free recall and discuss the implanted events using a computer online board. Doing so reduced false memories. The same participants were asked to discuss true events. Discussing true events via commenting on what their sibling wrote increased the amount of recall of those events and enhanced participants' tendency to recall the same event aspects. Prior to discussion, the photographs created false memories in 24% of people, but following discussion the rate of false memories decreased significantly. The authors noted that many participants were skeptical of the veracity of the false event and may have talked the siblings out of believing the false event.

In a second study, French, Garry, and Mori (2008) found a greater percentage of false memories following discussion among romantic partners than strangers, demonstrating that relationship status influences false memories. Other studies demonstrated that participants with no prior relationship could influence each other's memories via discussion. Garry et al. (2008) used the Mori technique, which allows participants to view the same movie at the same time with a twist: Some details in the movie differ for the two participants, who then discuss the movie. The results

showed that information that differed across the two participants that was accepted during the discussion was incorporated into memory.

Other researchers (Gabbert, Memon, & Wright, 2006) found that the person who spoke first in the discussion of a film was more influential in changing the partner's memory reports, regardless of whether or not they disagreed about what occurred. Additional never-encountered details in the film, rather than details related to what they viewed, were more likely to be incorporated into memory reports. Carlucci, Kieckhaefer, Schwartz, Villialba, and Wright (2011) showed that memory conformity can occur in the real world during a conversational encounter at the beach. After a confederate approached people, the experimenters asked people to identify him from a target-absent lineup. The first person to incorrectly pick a person out of a photo lineup influenced the responses of the second person. That is, in approximately 34% of the cases, the second person selected the same incorrect person, as did the first person. Clinical practitioners should be sensitive to the fact that discussing events can influence memories of the events recounted.

Hypnosis

The fact that hypnosis can inflate the plausibility or credibility of remembered events gives rise to concerns about the use of hypnosis for memory recovery. Such concerns are warranted by findings that hypnosis increases the sheer volume of recall, resulting in both more incorrect and correct information. When amount of recall is statistically controlled, hypnotic recall is no more accurate than nonhypnotic recall (see Erdelyi's [1994] narrative review of 34 studies and Steblay & Bothwell's [1994] meta-analysis of 24 studies) and results in increased confidence in memories reported, regardless of their accuracy (Lynn et al., in press). Additional findings indicate that (1) hypnosis produces more recall errors, more intrusions of uncued errors, and higher levels of memories for false information (Steblay & Bothwell, 1994); (2) enhanced recall confidence and false memories are associated with hypnotic responsiveness, although even relatively nonresponsive participants report false memories and unwarranted confidence in memories (Lynn, Myers, & Malinoski, 1997); (3) hypnosis is associated with an increase in the rate of early implausible memories prior to the accepted cutoff of infantile amnesia (age 2; Marmelstein & Lynn, 1999; Sivec, Lynn, & Malinoski, 1997a); (4) although hypnotized subjects are at least as likely as nonhypnotized participants to be misled in their recall by leading questions and sometimes exhibit recall deficits compared with nonhypnotized participants, there are indications that highly hypnotizable persons are particularly prone to memory errors in response to misleading information (see Spanos, 1996; Steblay & Bothwell, 1994); and (5) when participants are warned about possible memory problems associated with hypnotic recollections, they continue to report false memories during

and after hypnosis, although some studies indicate that warnings reduce the rate of pseudomemories in hypnotized and nonhypnotized individuals (Neuschatz, Lynn, Benoit, & Fite, 2002).

Some proponents of the use of hypnosis for memory recovery have argued that hypnosis is particularly helpful in facilitating recall of emotional or traumatic memories (Brown et al., 1998). However, contrary to this claim, eight studies (see Lynn, Lock, Myers, & Payne, 1997) that have compared hypnotic with nonhypnotic memory in the face of relatively emotionally arousing stimuli (e.g., films of shop accidents, depictions of fatal stabbings, a mock assassination, an actual murder videotaped serendipitously) have yielded two unambiguous conclusions: Hypnosis does not improve recall of emotionally arousing events, and arousal level does not moderate hypnotic recall. Krackow, Lynn, and Payne (2005–2006) found that hypnosis resulted in less consistent reports of an emotional event (memories of learning the news of Princess Diana's death) 11–12 weeks after learning the news compared with motivational instructions and context reinstatement/reverse order recall.

Although hypnosis is clearly contradicted for memory recovery, hypnosis does not necessarily yield more false memories than nonhypnotic procedures that are highly suggestive in nature (Lynn, Barnes, & Matthews, 2008). Indeed, any memory recovery procedure that is suggestive in nature or conveys the expectation that accurate memories can easily be recovered is likely to increase the sheer volume of memories and to bolster confidence in inaccurate as well as accurate memories. However, simply asking participants to focus on the task at hand and to do their best to recall specific events yield accurate recall comparable to hypnosis, but with fewer or comparable recall errors (Krackow et al., 2005–2006; see also Lynn, Lock, Myers, & Payne, 1997).

Nonhypnotic and Hypnotic Age Regression

Age regression involves "regressing" a person back through time to an earlier life period. Subjects are typically given relaxation instructions and then asked to mentally re-create events that occurred at successively earlier periods in the person's life, or to focus on a particular event at a specific age, with suggestions that they are to fully relive the target event. A televised documentary (Bikel, 1995) showed a group therapy session in which a woman was age regressed through childhood to the womb and eventually to being trapped in her mother's Fallopian tube. This case is not an isolated example of the kind of experience we consider extremely questionable. The mental health and counseling literature contains numerous case studies of prenatal remembering that are nearly, if not equally, unlikely (van Husen, 1988; Lawson, 1984).

The literature strongly suggests that the experiences of age-regressed individuals are contextually dependent and expectancy-driven social

constructions: Age-regressed participants behave according to cues they derive from the social situation, and their knowledge and beliefs about age-relevant behaviors reflect their fantasies and beliefs and assumptions about childhood, rather than being literal reinstatements of childhood experiences, behaviors, and feelings.

Hypnosis is often used to facilitate the experience of age regression. However, Nash's (1987) review revealed that no special correspondence exists between the behaviors and experiences of hypnotized adults and those of actual children. For example, adult participants age regressed to childhood do not perform as expected on Piagetian (e.g., conservation) tasks. In fact, as the following two studies indicate, hypnotic age regression can contribute to memory distortions of early-life events.

Nash, Drake, Wiley, Khalsa, and Lynn (1986) attempted to corroborate the memories of subjects who had participated in an earlier age regression experiment. In this study, hypnotized and role-playing (i.e., simulating) participants were regressed to age 3 to a scene in which they were in the soothing presence of their mothers. During the experiment, subjects reported the identity of their transitional objects (e.g., blankets, teddy bears). Third-party verification (parent report) of the accuracy of recall regarding the transitional object was obtained for 14 hypnotized subjects and 10 simulation control subjects. Despite the similarity to children in their means of relating to transitional objects, hypnotic subjects were less able than were control subjects to correctly identify the specific transitional objects actually used. Hypnotic subjects' hypnotic recollections, for example, matched their parents' reports only 21% of the time. In contrast, the parents of simulators' corroborated their reports 70% of the time. All recollections obtained during hypnosis were incorporated into hypnotic recollections, regardless of accuracy.

Sivec, Lynn, and Malinoski (1997b) age regressed participants to the age of 5 and suggested that girls play with a Cabbage Patch doll and boys with a He-Man toy. An important aspect of this study was that these toys were not released until 2 or 3 years after the target time of the age regression. Half of the subjects received hypnotic age regression instructions, and the other half received suggestions to age regress that were not administered in a hypnotic context. Interestingly, none of the nonhypnotized persons was influenced by the suggestion. In contrast, 20% of the hypnotized subjects rated the memory of the experience as real and were confident that the event occurred at the age to which they were regressed. In sum, the findings strongly contradict the use of age regression techniques in psychotherapy.

Past-Life Regression

The search for traumatic memories can extend to before birth (see Mills & Lynn, 2000). One type of therapy, called "past-life regression therapy," is based on the premise that traumas that occurred in previous lives influence current psychological and physical symptoms (e.g., Woolger, 1988). For

example, Weiss (1988) published a widely publicized series of cases focusing on patients who were hypnotized and age regressed to "go back to" the source or origin of a particular present-day problem. When the patients were regressed, they reported events that Weiss and his patients interpreted as having their source in previous lives.

When regressed persons have seemingly realistic and detailed experiences during regression, they may seem very convincing to both patient and therapist. However, when Spanos, Menary, Gabora, DuBreuil, and Dewhirst (1991) examined past-life reports for historical accuracy, the information participants provided about the specific time period during their hypnotic age regression enactments was almost "invariably incorrect" (p. 137). For example, one participant who was regressed to ancient times claimed to be Julius Caesar, emperor of Rome, in 50 B.C., even though the designations of B.C. and A.D. were not adopted until centuries later and even though Julius Caesar died decades prior to the first Roman emperor.

Additionally, it is possible to elicit and manipulate past-life reports by structuring participants' expectancies. For example, in one study (Spanos et al., 1991, Study 2), participants were informed at the outset that past-life identities were likely to be of a different gender, culture, and race from that of the present personality. In contrast, other participants received no prehypnotic information about their past-life identities. Spanos and his colleagues (1991) found that participants' past-life experiences were not only quite elaborate but that they tended to conform to induced expectancies. In another study, Spanos and his associates (Spanos et al., 1991, Study 3) showed that participants' past-life reports during hypnotic age regression varied in terms of the prehypnotic information they received about whether children were frequently abused during past historical periods. The researchers also found that past-life reports were associated with participants' prior beliefs in reincarnation. Interestingly, Meyersberg, Bogdan, Gallo, and McNally (2009) found that individuals who report recovered memories of past lives exhibit significantly higher false recall and recognition rates with respect to a false memory task (Desse–Roediger–McDermott paradigm), compared with participants who report having lived only one life.

Pyun and Kim (2009) analyzed memories obtained using past-life regression in combination with hypnosis in Korean and Canadian cultures. The content of the memories was influenced by culture and religion, with past-life identities of humans versus animals varying as a function of participant culture. Spanos (1996) concluded that hypnotically induced past-life experiences are rule-governed, goal-directed fantasies that are context generated and sensitive to the demands of the hypnotic regression situation. Such imaginative scenarios are constructed from available cultural narratives about past lives and known or surmised details and facts regarding specific historical periods. In short, age regression techniques, with or without accompanying hypnosis, are highly unreliable methods of enhancing recall of early life experiences, and individuals who report such experiences spontaneously may be at special risk for forming false memory.

Bibliotherapy and Abuse Checklists

Therapists working with suspected abuse victims often inform them that the symptoms they experience are suggestive of a history of abuse (Bass & Davis, 1988; Blume, 1990; Frederickson, 1992; Loftus, 1993), followed by an explanation that abuse memories can be recovered because memory functions like a tape or video recorder. Treatment then focuses on recovering those memories for the purposes of healing past psychological wounds.

Many therapists who treat patients with suspected abuse histories prescribe "survivor books" or self-help books (see also Rosen, Glasgow, Moore, & Barrera, Chapter 9, this volume) to provide "confirmation" that the individual's symptoms are due to past abuse and to provide a means of gaining access to memories. The books typically provide imaginative exercises and stories of other survivors' struggles (Lindsay & Read, 1994). Some of the most influential popular books of this genre include Bass and Davis's (1988) *The Courage to Heal*, Frederickson's (1992) *Repressed Memories*, and Blume's *Secret Survivors: Uncovering Incest and Its Aftereffects in Women* (1990). These books offer potential support for, and validation to, actual abuse survivors. However, the fact that the writers interpret current psychological symptoms as indicative of an abuse history, invite readers to imaginatively review their past experiences, and include suggestive stories of survivors of abuse may increase the risk that readers incorporate false memories of abuse into their archive of personal memories.

Mazzoni, Loftus, and Kirsch (2001) provided a dramatic illustration of how reading material and psychological symptom interpretation can increase the plausibility of an initially implausible memory of witnessing a demonic possession. The authors conducted their study in Italy, where demonic possession is viewed as a more plausible occurrence than in the United States. However, in an initial testing session, all of the participants indicated that demonic possession was not only implausible, but that it was very unlikely that they had witnessed an occurrence of possession as children.

A month after the first session, participants in one group first read three short articles (in a pack of 12), which indicated that demonic possession is more common than is generally believed and that many children have witnessed such an event. These participants were compared with (1) participants who read three short articles about choking and (2) control participants who received no manipulation. Individuals who received one of the manipulations returned to the laboratory the following week and, based on their responses to a fear questionnaire they completed, were informed (regardless of their actual responses) that their fear profile indicated that they had probably either witnessed a possession or had almost choked during early childhood.

When the students returned to the laboratory for a final session and completed the original questionnaires, they indicated that the two suggested events—witnessed possession and choking—were more real than

before. Additionally, 18% of the participants indicated that they had probably witnessed a possession. No changes in memories were evident in the control condition that read articles about choking. Clearly, therapists should exercise caution in implying that such events as abuse might have occurred on the basis of current symptoms alone.

Many popular psychology self-help books on the topic of incest include lists of psychological symptoms (e.g., "Do you have trouble knowing what you want?" "Do you use work or achievements to compensate for inadequate feelings in other parts of your life?") that are presented as possible or probable correlates of childhood incest. Blume's (1990) Incest Survivors' Aftereffects Checklist consists of 34 such correlates. The scale instructions read: "Do you find many characteristics of yourself on this list? If so, you could be a survivor of incest." Blume also indicates that "clusters" of these items are significant predictors of childhood sexual abuse and that "the more items endorsed by an individual the more likely that there is a history of incest" (pp. xviii–xxi).

Many of the characteristics on such checklists are quite vague and seemingly applicable to many nonabused individuals. Much of the seeming "accuracy" of such checklists could stem from the Barnum effect—the tendency to believe that highly general statements that are true of many individuals in the population apply specifically to oneself (see Garb & Boyle, Chapter 2, this volume, for a more detailed discussion of the Barnum effect). Indeed, Emery and Lilienfeld (2004) found that a checklist whose items were drawn from widely used child sexual abuse survivor descriptions in the popular psychology literature (e.g., "I daydream frequently," "I often give too much to others in relationships") correlated highly with a checklist of Barnum items (e.g., "I at times worry too much," "I sometimes feel insecure around people I've just met"). In addition, this checklist failed to distinguish abused from nonabused women when global psychopathology was statistically controlled. Moreover, there is no known constellation of specific symptoms that is indicative of a history of child sexual abuse (Beitchman et al., 1992). Some genuine victims of childhood incest may experience many symptoms, others only some, and still others none. Moreover, nonvictims experience many of the same symptoms often associated with sexual abuse (Tavris, 1993). Accordingly, abuse checklists and interpreting current symptoms as indicators of abuse are unreliable and should be eschewed.

INDIVIDUAL DIFFERENCES

The General Public: Popular Beliefs

Laypeople hold misconceptions about how human memory works. Surveys have shown that many people believe in the permanence of human memory (Garry, Loftus, & Brown, 1994; Loftus & Loftus, 1980) and in the ability of techniques such as hypnosis to assist in memory recovery (McConkey

& Jupp, 1986). Additionally, a substantial proportion of subjects in Garry and colleagues' (1994) research believed in the veridicality of memories retrieved from the womb, and Yapko (1994) and Garry and colleagues (Garry et al., 1994) found that a substantial proportion of lay samples believed in the existence of past lives. More recently, Simons and Chabris (2011, 2012) reported that 63% of the U.S. public agreed that memory works like a video camera, 48% agreed that memory is permanent, and 55% believed memory can be enhanced through hypnosis.

Practitioners

Practitioners, like laypeople, often hold inaccurate beliefs about memory. Yapko (1994) found that 47% of professionals (the majority of whom were practitioners) had greater faith in the veridicality of hypnotic than non-hypnotic memories, and 31% believed that events recalled during hypnosis were likely to be accurate. Furthermore, 54% of his sample believed to some degree in the effectiveness of hypnosis for recovering memories as far back as birth, and 28% believed in its effectiveness for recovering memories from past lives. Magnussen and Melinder (2012) surveyed psychologists recruited from the Norwegian Psychological Association and found that their level of knowledge did not exceed that of laypersons or trial judges with respect to beliefs about eyewitness memory. Moreover, a substantial minority of those surveyed held scientifically unsubstantiated ideas about memory (e.g., children's recall is typically as good as adult recall); 63% of those surveyed indicated that most recovered memories are real, and 38% believed that adults can repress memories of traumatic events like committing murder. In a recent study conducted in the United Kingdom (Ost, Wright, Easton, Hope, & French, 2013), the belief in the accuracy of satanic/ritual abuse memories and the diagnosis of dissociative identity disorder correlated negatively with the belief that false memories were possible.

Individual Differences in Susceptibility to False Memories

Research has provided evidence for small to moderate, but consistent, relations between dissociation and recall errors (see Eisen & Lynn, 2001, for a review). In addition to dissociation, vividness of visualization has been found to be related to increased suggestibility and/or false memories and source monitoring difficulties (see Eisen & Lynn, 2001; Ost, Granhag, Udell, & Hjelmsater, 2008; Porter, Taylor, & ten Brinke, 2009), as have absorption and magical ideation (McNally et al., 2004), fantasy proneness (Geraerts, Smeets, Jelicic, van Geerden, & Merckelbach, 2005), compliance (Malinoski & Lynn, 1999), schizotypy (Clancy, McNally, Schacter, Lenzenweger, & Pitman, 2002), divided attention (in adults but not children; Otgaar, Peters, & Howe, 2012), hypnotizability (see Lynn, Boycheva,

Deming, Lilienfeld, & Hallquist, 2009), interrogative suggestibility (Malinoski & Lynn, 1999), and the combination of high extraversion in interviewers and low extraversion in interviewees (Porter, Birt, Yuille, & Lehman, 2000). Interestingly, people with higher cognitive ability as measured by standard tests of intelligence are somewhat more resistant to false memory implantation (Zhu et al., 2010). Clancy, Schacter, McNally, and Pitman (2000) found that women who reported recovered and repressed memories of childhood sexual abuse are more prone to false recognition errors, compared with women with no history of childhood sexual abuse and women who were sexually abused as children but who always remembered the abuse. More recently, Geraerts et al. (2005) found that people who recovered memories of abuse through suggestive therapy were more vulnerable to false memories. The available evidence implies that clinicians should be especially vigilant to avoid leading interview procedures with clients who evidence higher levels of the characteristics cited above.

CRITICISMS OF THE LITERATURE

There have been three major criticisms of the research we have reviewed. First, studies that rely on parents of siblings as a means of implanting memories have been criticized (Freyd, 1998; Pope, 1996; Pezdek & Lam, 2007; Pope & Brown, 1996) as inadequate analogues of what occurs in the "real world" and psychotherapy. Accordingly, critics suggest that studies such as the Lost-in-the-Mall study (i.e., Loftus & Pickrell, 1995) demonstrate the suggestive influence of siblings or parents who claim to be present at the time of the event, rather than the influence of therapists who were not present at the time of the event. However, this criticism does not apply to the numerous studies we reviewed that did not use parents' or siblings' reports to lend credence to a false memory that was elicited, but demonstrated that memories for implausible or impossible events (e.g., witnessing demonic possession) could be instated with memory recovery techniques.

A second criticism concerns whether the implanted memory is false. Brown and colleagues (1998) claimed that the experimenter can never know if the target false event actually occurred; therefore, the "implanted false memory" may be real. However, in many studies we reviewed, the experimenter controls the stimuli presented to participants and is fully aware of what events are "real" versus "fabricated."

Brown and colleagues (1998) also suggested that parents' memory reports do not necessarily represent more accurate accounts of the past than participants' renditions of what occurred. Accordingly, the failure to corroborate a particular memory based on parental report provides no guarantee that the event the participant remembers did not occur. However, these criticisms do not apply to the studies we reviewed that targeted impossible or unlikely events with a vanishingly small probability (e.g.,

UFO abduction) of having occurred in real life or events recovered below the cutoff or infantile amnesia.

The third critique of the memory implantation literature is that it is difficult if not impossible to generalize laboratory research to reports of abuse in clinical settings (Pope, 1996; Pezdek & Lam, 2007; Pope & Brown, 1996). Critics contend that a memory of being abused cannot be easily implanted. That is, abuse is a complex event that typically occurs repeatedly and would therefore be more difficult to implant (cf. Brewin, 1997; Brown et al., 1998; Olio, 1994). The critics do not offer a clear explanation of why this would be the case. Nevertheless, although they raise an interesting question (i.e., Are complex repeated events more difficult to implant than relatively less complex, single occurrences?) that should be explored in future research, there is by now a substantial body of evidence indicating that complex autobiographical memories can be implanted in studies using ecologically valid populations or materials (see Wade et al., 2007).

Critics also contend that laboratory research may not generalize to the real world because the events that are typically suggested do not approximate the degree of trauma experienced by actual survivors of childhood sexual abuse. Although critics assume that a highly traumatic memory such as abuse would be much more difficult to implant (Olio, 1994; Pope & Brown, 1996) than a less traumatic memory, they do not advance a clear rationale for this contention. Moreover, previous research shows that people can develop posttraumatic stress symptoms of highly implausible events that likely represent false memories (McNally et al., 2004), thereby decreasing the credibility of this criticism.

The critics of the research we have reviewed have correctly identified generalizability as an important issue. It is difficult to make informed comparisons regarding false memory rates in clinical versus laboratory contexts because the precise rates of false memories associated with the procedures we have reviewed in clinical situations are unknown. In clinical situations, little is known about the rates of (1) false memories that arise spontaneously due to the vagaries of ordinary memory, (2) false memories that are elicited by memory recovery procedures, (3) accurate memories that arise spontaneously in the process of therapeutic discourse, and (4) accurate memories that are elicited by memory recovery procedures.

It seems plausible, however, if not likely, that expectancies, suggestive procedures, and demand characteristics play a far more significant role in therapeutic situations than in experimental contexts. The therapist's potential to exert social influence on a help-seeking, eager-to-please, vulnerable patient over many months or years is likely much greater than the experimenter's influence on a subject participating in a "one-shot" experiment for money or course credit. Accordingly, the effects of hypnosis, guided imagery, suggestion, and symptom interpretation on memory may be more, rather than less, pronounced in a clinical than in a laboratory context.

HYPOTHESIZED PATH OF FALSE MEMORY CREATION

We believe that imaginative narratives of sexual abuse that never occurred, past life reports, and alien abduction experiences can and do arise in the context of a treatment in which the patient comes to believe that the narrative provides a plausible explanation for current life difficulties. The narrative can achieve a high degree of plausibility due to a variety of factors we have delineated. The most important of these factors include the following:

1. Cultural beliefs that link the idea of abuse with psychopathology and the repression of traumatic memories.
2. The therapist's support or suggestion of this interpretation.
3. The failure to consider alternative explanations for problems in living presumed to be associated with abuse.
4. The search for confirmatory data (see also Garb & Boyle, Chapter 2, this volume).
5. Repeated unsuccessful attempts to recall past events leading to the perception of significant "amnesia" for the past and justifying the use of memory recovery procedures (Belli et al., 1998). Normalizing the occurrence of amnesia or repression associated with traumatic events enhances the plausibility of events recalled, regardless of their accuracy (Scoboria, Lynn, Hessen, & Fisico, 2007).
6. The use of suggestive memory recovery techniques that further increase the plausibility of abuse (Pezdek, Finger, & Hodge, 1997) and yield information and remembrances consistent with the idea that abuse occurred.
7. Memory recovery techniques that engender source monitoring errors; the client adopts the suggested or imagined content as historical fact.
8. Physiological arousal that occurs as the client discusses or recalls the false event, which enhances the credibility of the recollection and the perception that the event actually occurred (Krackow & Rabenhorst, 2010).
9. Increasing commitment to the narrative on the part of the client and therapist, escalating dependence on the therapist, and anxiety reduction associated with ambiguity reduction.
10. Encouragement of a "conversion" or "coming out" experience by the therapist or supportive community (e.g., therapy group), which solidifies the role associated with the narrative (e.g., "abuse victim") and bolsters feelings of empowerment, which constitute positive reinforcement.
11. The fact that the narrative provides continuity of the past and the future, along with comfort, belonging, and identity.

SHOULD THERAPISTS ENGAGE IN MEMORY RECOVERY?

The extant evidence provides little support for the use of memory recovery techniques in uncovering memories of abuse in psychotherapy. In fact, there are a number of reasons for therapists to eschew these procedures.

1. "Recovered memory therapy . . . is predicated on the trauma memory argument—that memories of traumatic events have special properties that distinguish them from ordinary memories of the sort usually studied in the laboratory" (Shobe & Kihlstrom, 1997, p. 70). "Nothing about the clinical evidence suggests that traumatic memories are less subject to reconstructive processes or that special techniques are required to recover them" (Shobe & Kihlstrom, 1997, p. 74; but see Nadel & Jacobs, 1998, for a different view).

2. Most survivors of traumatic abuse past the age of 3 do not forget their abuse. In fact, the literature points to the opposite conclusion: In general, memory following traumatic events is enhanced compared with memory following nontraumatic events (see McNally, 2007; Shobe & Kihlstrom, 1997).

3. Even if a certain percentage of accurate memories can be recovered in therapy, it may be unprofitable to do so. As Lindsay (1996) observed, "numerous lines of evidence suggest that only a very small percentage of psychotherapy clients have problems that are caused by non-remembered histories of abuse" (p. 64). Certainly, no empirical work reveals a casual connection between nonremembered abuse and psychopathology.

4. There is no demonstrable benefit associated with the straightforward catharsis of emotional events in treatment. To the contrary, the mere experience and expression of painful memories and emotions, when not grounded in attempts to engender positive coping and mastery, can be harmful (Berkowitz, 2012; Littrell, 1998).

5. There is no empirical basis for arguing that hypnotic or nonhypnotic memory recovery procedures are more effective than present-centered approaches in treating the repercussions of trauma or psychological problems in general. Indeed, there is no empirically supported psychotherapy or procedure that relies on the recovery of forgotten traumatic events to achieve a positive therapeutic outcome (Chambless & Ollendick, 2001). Adshead (1997) has gone so far as to argue that if memory work with trauma patients is not effective, then "it would therefore be just as unethical to use memory work for patients who could not use it or benefit by it, as it would be to prescribe the wrong medication, or employ useless surgical technique" (p. 437).

We further contend that if therapists decide to use memory recovery techniques, then they should provide their clients with a written informed consent document that apprises the clients of (1) accurate, scientifically grounded information about the reconstructive nature of memory; (2) the fact that recovered memories must be corroborated before they can be given

special credence; and (3) information regarding laboratory studies of memory pertinent to the techniques employed.

CONCLUSIONS

Let us be clear about what our findings do not mean as well as what they do mean. First, our findings do not imply that all memory recovery techniques are necessarily problematic. For example, the "revised cognitive interview" (Fisher & Geiselman, 1992), which incorporates basic techniques derived from experimental research on memory (e.g., providing subjects with appropriate retrieval cues, searching for additional memorial details), has garnered empirical support as a method of enhancing memory in eyewitness contexts (Memom, Meissner, & Fraser, 2010). Some of the techniques comprising this interview could ultimately prove helpful in the therapeutic context for enhancing memories of specific events.

Second, we do not wish to imply that all uses of hypnosis in psychotherapy are problematic. To the contrary, some controlled research suggests that hypnosis may be a useful adjunct to cognitive-behavioral therapy, pain control procedures, obesity treatment, and smoking cessation treatments (Lynn, Rhue, & Kirsch, 2010), although the extent to which hypnosis provides benefits above and beyond relaxation (and other nonspecific effects) in such cases remains unclear.

Third, we do not wish to claim that all memories recovered after years or decades of forgetting are necessarily false. Several case reports are consistent with the possibility that people can sometimes recall old childhood experiences after long periods of nonrecall (Geraerts et al., 2009; Schooler, Ambadar, & Bendiksen, 1997; but see Loftus & Guyer, 2002). In some of these cases, there is evidence that such apparently unrecalled memories were previously recalled (see Geraerts et al., 2009). Nevertheless, we remain open to the possibility that certain recovered childhood memories are veridical, although we believe that further research will be needed to document their existence and possible prevalence. These important and unresolved issues notwithstanding, the conclusion that certain suggestive therapeutic practices, particularly those that we have discussed in this chapter, can foster false memories in some clients appears indisputable. We urge practitioners to exercise considerable caution when using these techniques in psychotherapy and to base their memory-related therapeutic practices on the best available scientific evidence.

GLOSSARY

DRM paradigm: In the Deese–Roediger–McDermott (DRM; Roediger & McDermott; see also Deese, 1959) paradigm, participants are presented (auditorily) with a series of lists of thematically related words such as bed, rest, wake, doze, dream,

and pillow, all of which are thematically related to such nonpresented words as sleep. They then are administered an oral recognition test consisting of the words actually presented in the experiment, "critical" nonpresented words that are thematically related to the presented words (e.g., sleep), and a series of nonpresented, thematically related, words. Using this paradigm, many participants (approximately 70%, on average) experience "memory illusions"; that is, they falsely recognize thematically related nonpresented words with a high degree of confidence.

Hypnosis: The American Psychological Association, Division of Psychological Hypnosis, has adopted a consensus definition of hypnosis as a procedure during which a health professional or researcher suggests that a client, patient, or subject experience changes in sensations, perceptions, thoughts, or behavior. The hypnotic context is generally established by an induction procedure. Although there are many different hypnotic inductions, most include suggestions for relaxation, calmness, and well-being.

Hypnotizability: The degree or extent of responsiveness to suggestions administered in a situation that is defined as hypnosis. How people respond to hypnosis is dependent largely on their motivation to respond; their beliefs, attitudes, and expectancies about hypnosis and how they will respond; and their responsiveness to waking imaginative suggestions.

Infantile amnesia: Virtually all contemporary memory researchers agree that accurate memory reports of events that occur before 24 months of age are extremely rare. This inability to recall very early life events is known as "infantile amnesia," a phenomenon attributable to developmental changes that influence how children process, retrieve, and share information (see Malinoski et al., 1998).

Interrogative suggestibility: Interrogative suggestibility involves the tendency of an individual's account of events to be altered by misleading information and interpersonal pressure within an interview. The Gudjonsson Scale of Interrogative Suggestibility (GSS; Gudjonsson, 1984) was initially used to predict individual differences in susceptibility to highly suggestive and misleading questioning during a police interrogation, but it has been used widely in research on memory and suggestibility.

REFERENCES

Adshead, G. (1997). *Written on the body: Deliberate self-harm and violence. A practical guide to forensic psychotherapy.* London: Kingsley.

Austin, S. (1995). *Exploration of a conditioning model in understanding the manifestation of physical symptoms of sexual abuse survivors.* Unpublished doctoral dissertation, University of Denver, CO.

Bass, E., & Davis, L. (1988). *The courage to heal.* New York: Harper & Row.

Beitchman, J. H., Zucker, K. J., Hood, J. E., DaCosta, G. A., Akman, D., & Cassavia E. (1992). A review of the long-term effects of child sexual abuse. *Child Abuse and Neglect, 16*(1), 101–118.

Belli, R. F., Winkielman, P., Read, J. D., Schwarz, N., & Lynn, S. J. (1998). Recalling more childhood events leads to judgments of poorer memory: Implications for the recovered/false memory debate. *Psychonomic Bulletin and Review, 5,* 318–323.

Berkowitz, L. (2012). A different view of anger: The cognitive-neoassociation conception of the relation of anger to aggression. *Aggressive Behavior, 38*(4), 322–333.

Berkowitz, S. R., Laney, C., Morris, E. K., Garry, M., & Loftus, E. F. (2012). Pluto behaving badly: False beliefs and their consequences. *American Journal of Psychology, 121*(4), 643–660.

Bernstein, D. M., Laney, C., Morris, C. K., & Loftus E. F. (2005). False memories about food can lead to avoidance [Special issue]. *Social Cognition, 23,* 11–34.

Bernstein, D. M., & Loftus, E. F. (2009). The consequences of false memories for food preferences and choices. *Perspectives on Psychological Science, 4*(2), 135–139.

Bikel, O. (Producer). (1995). Divided memories. In *Frontline*. New York: CBS.

Bindler, J. L., & Smokler, I. (1980). Early memories: A technical aid to focusing in time-limited dynamic psychotherapy. *Psychotherapy: Theory, Research and Practice, 17,* 52–62.

Blume, E. S. (1990). *Secret survivors: Uncovering incest and its aftereffects in women*. New York: Wiley.

Brewin, C. R. (1997). Commentary on dispatch from the (un)civil memory wars. In D. Read & S. Lindsay (Eds.), *Recollections of trauma: Scientific research and clinical practice* (pp. 194–196). New York: Plenum Press.

Brown, D., Scheflin, A. W., & Hammond, D. C. (1998). *Memory, trauma treatment, and the law*. New York: Norton.

Bruhn, A. R. (1984). The use of earliest memories as a projective technique. In P. McReynolds & C. J. Chelune (Eds.), *Advances in psychological assessment* (Vol. 6, pp. 109–150). San Francisco: Jossey-Bass.

Carlucci, M. E., Kieckhaefer, J. M., Schwartz, S. L., Villalba, D. K., & Wright, D. B. (2011). The South Beach study: Bystanders' memories are more malleable. *Applied Cognitive Psychology, 25,* 562–566.

Chan, J. C. K., & La Paglia, J. A. (2011). The dark side of atesting memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied*, *17*(4), 418–432.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52,* 685–716.

Clancy, S. A., McNally, R. J., Shacter, D. L., Lenzenweger, M. F., & Pitman, R. K. (2002). Memory distortion in people reporting abduction by aliens. *Journal of Abnormal Psychology, 111,* 455–461.

Clancy, S. A., Schacter, D. L., McNally, R. J., & Pitman, R. K. (2000). False recognition in women reporting recovered memories of sexual abuse. *Psychological Science, 11*(1), 26–31.

DuBreuil, S. C., Garry, M., & Loftus, E. F. (1998). Tales from the crib: Age-regression and the creation of unlikely memories. In S. J. Lynn & K. M. McConkey (Eds.), *Truth in memory* (pp. 137–160). Washington, DC: American Psychological Association.

Eisen, M., & Lynn, S. J. (2001). Memory, suggestibility, and dissociation in children and adults. *Applied Cognitive Psychology, 15,* 49–73.

Emery, C. L., & Lilienfeld, S. O. (2004). The validity of childhood sexual abuse checklists in the popular psychology literature: A Barnum effect? *Professional Psychology: Research and Practice*, *35*, 268–274.

Erdelyi, M. (1994). Hypnotic hypermnesia: The empty set of hypermnesia. *International Journal of Clinical and Experimental Hypnosis, 42,* 379–390.

Farmer, S. (1989). *Adult children of abusive parents: A healing program for those who have been physically, sexually, or emotionally abused*. Los Angeles: Lowell House.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, IL: Charles C Thomas.

Fox, S. A. (2002). *A survey of mental health clinicians' use of dream interpretation in psychotherapy* (doctoral dissertation). Ann Arbor, MI: ProQuest Information and Learning.

Frederickson, R. (1992). *Repressed memories*. New York: Fireside/Parkside.

French, L., Garry, M., & Mori, K. (2008). You say tomato?: Collaborative remembering leads to more false memories for intimate couples than for strangers. *Memory*, *16*, 262–273.

French, L., Sutherland, R., & Garry, M. (2006). Discussion affects memory for true and false childhood events. *Applied Cognitive Psychology, 20,* 671–680.

Freud, S. (1953). The interpretation of dreams. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth Press. (Original work published 1900)

Freud, S. (1955). From the history of an infantile neurosis. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth Press. (Original work published 1918)

Freyd, J. J. (1998). Science in the memory debate. *Ethics and Behavior, 8*(2), 101–113.

Gabbert, F., Memon, A., & Wright, D. B. (2006). Memory conformity: Disentangling the steps toward influence during a discussion. *Psychonomic Bulletin and Review, 13*(3), 480–485.

Garry, M., French, L., Kinzett, T., & Mori, K. (2008). Eyewitness memory following discussion: Using the MORI technique with a western sample. *Applied Cognitive Psychology, 22*, 431–439.

Garry, M., Loftus, E. F., & Brown, S. W. (1994). Memory: A river runs through it [Special issue]: The recovered memory/false memory debate. *Consciousness and Cognition: An International Journal, 3,* 438–451.

Garry, M., Manning, C., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin and Review, 3,* 208–214.

Garry, M., Sherman, S. J., Wade, K. A., Hunt, M. J., & Smith, P. J. (2001). Imagination inflation is a fact, not an artifact. *Memory and Cognition, 29*(5), 719–729.

Geraerts, E., Bernstein, D. M., Merckelbach, H., Linders, C., Raymackers, L., & Loftus, E. F. (2008). Lasting false beliefs and their behavioral consequences. *Psychological Science, 19*(8), 749–753.

Geraerts, E., Lindsay, D. S., Merckelbach, H., Jelicic, M., Raymaekers, L., Arnold, M. M., & Schooler, J. W. (2009). Cognitive mechanisms underlying recovered-memory experiences of childhood sexual abuse. *Psychological Science*, *20*(1), 92–98.

Geraerts, E., Smeets, E., Jelicic, M., van Geerden, J., & Merckelbach, H. (2005). Fantasy proneness, but not self-reported trauma is related to DRM performance of women reporting recovered memories of childhood sexual abuse. *Consciousness and Cognition, 14,* 602–612.

Gerrie, M. P., Belcher, L. E., & Garry, M. (2006). Mind the gap: False memories for missing aspects of an event. *Applied Cognitive Psychology*, *20*(5), 689–696.

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences*, *5*(3), 303–314.

Heaps, C., & Nash, M. (1999). Individual differences in imagination inflation. *Psychonomic Bulletin and Review, 6*(2), 313–318.

Howe, M. L. (2003). Memories from the cradle. *Current Directions in Psychological Science, 12*(2), 62–65.

Hyman, I. E., Jr., Husband, T. H., & Billings, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology, 9,* 181–197.

Hyman, I. E., Jr., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language, 35*, 101–117.

Krackow, E. (2010). Narratives distinguish experienced from imagined childhood events. *American Journal of Psychology, 123*(1), 71–80.

Krackow, E., Lynn, S. J., & Payne, D. G. (2005–2006). The death of Princess Diana: The effects of memory enhancement procedures on flashbulb memories. *Imagination, Cognition and Personality, 25*(3), 197–219.

Krackow, E., & Rabenhorst, M. (2010). Does the body know best?: Psychophysiological reactivity to perceived versus imagined childhood events. *Imagination, Cognition and Personality, 30*(2), 133–145.

Lampinen, J. M., Odegard, T. N., & Bullington, J. L. (2003). Qualities of memories for performed and imagined actions. *Applied Cognitive Psychology, 17,* 881–893.

Laney, C., & Takarangi, M. K. T. (2013). False memories for aggressive acts. *Acta Psychologica, 14*(2), 227–234.

Lawson, A. J. (1984). Perinatal imagery in UFO abduction reports. *Journal of Psychohistory, 12,* 211–239.

Legault, E., & Laurence, J. R. (2007). Recalled memories of child sexual abuse: Social work, psychology and psychiatric reports of beliefs, practices, and cases. *Australian Journal of Experimental and Clinical Hypothesis, 35*(2), 111–133.

Levis, D. J. (1995). Decoding traumatic memory: Implosive theory of psychopathology. In W. O'Donohue & L. Krasner (Eds.), *Theories in behavior therapy* (pp. 173–207). Washington DC: American Psychological Association.

Lindsay, D. S. (1996). Conceptualizing and clarifying criticisms of memory work in psychotherapy. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate* (pp. 267–278). San Diego, CA: Academic Press.

Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science, 15*(3), 149–154.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory and Cognition, 17,* 349–358.

Lindsay, D. S., & Read, D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology, 8,* 281–338.

Littrell, J. (1998). Is the experience of painful emotion therapeutic? *Clinical Psychology Review, 18,* 71–102.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

Loftus, E. F., & Guyer, M. (2002). Who abused Jane Doe?: The hazards of the single case history. *Skeptical Inquirer, 26*(3), 24–32.

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memories*. New York: Plenum.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the brain. *American Psychologist, 35,* 409–420.

Loftus, E. F., & Mazzoni, G. (1998). Using imagination and personalized suggestion to change behavior. *Behavior Therapy, 29,* 691–708.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25,* 720–725.

Lynn, S. J., Barnes, S. M., & Matthews, A. (2008). Hypnosis and memory: From Bernheim to the present. In K. D. Markman, W. M. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental stimulation* (pp. 103–118). New York: Psychology Press.

Lynn, S. J., Boycheva, E., Demling, A., Lilienfeld, S. O., & Hallquist, M. N. (2009). Forensic hypnosis: The state of science. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 80–99). New York: Guilford Press.

Lynn, S. J., Lemons, P., Green, J. P, Mazzoni, G., Lilienfeld, S. O., & Kisch, I. (in press). Forensic hypnosis. In R. Cautin & S. O. Lilienfeld (Eds.), *Encyclopedia of clinical psychology*. New York: Wiley.

Lynn, S. J., Lock, T. G., Myers, B., & Payne, D. G. (1997). Recalling the unrecallable: Should hypnosis be used to recover memories in psychotherapy? *Current Directions in Psychological Science, 6,* 79–83.

Lynn, S. J., Malinoski, P., & Green, J. (1999). *Early memory reports as a function of high versus low expectancy*. Unpublished manuscript, State University of New York at Binghamton.

Lynn, S. J., Malinoski, P., & Sivec, H. (1998). The assessment, validity, and determinants of early memory reports: A critical review. In S. J. Lynn & K. M. McConkey (Eds.), *Truth in memory* (pp. 109–136). New York: Guilford Press.

Lynn, S. J., Myers, B., & Malinoski, P. (1997). Hypnosis, pseudomemories, and clinical guidelines: A sociocognitive perspective. In D. Read & S. Lindsay (Eds.), *Recollections of trauma: Scientific research and clinical practice* (pp. 305–331). New York: Plenum Press.

Lynn, S. J., Neuschatz, J., Fite, R., & Rhue, J. R. (2002). Hypnosis and memory: Implications for the courtroom and psychotherapy.In M. Eisen & G. Goodman (Eds.), *Memory, suggestion, and the forensic interview* (pp. 287–307). New York: Guilford Press.

Lynn, S. J., Rhue, J. W., & Kirsch, I. (Eds.). (2010). *Handbook of clinical hypnosis* (2nd ed.). Washington, DC: American Psychological Association.

Magnussen, S., & Melinder, A. (2012). What psychologists know and believe about memory: A survey of practitioners. *Applied Cognitive Psychology, 26*(1), 54–60.

Malinoski, P., & Lynn, S. J. (1999). The plasticity of very early memory reports: Social pressure, hypnotizability, compliance, and interrogative suggestibility. *International Journal of Clinical and Experimental Hypnosis*, *47*, 320–345.

Malinoski, P., Lynn, S. J., & Sivec, H. (1998). The assessment, validity, and determinants of early memory reports: A critical review. In S. J. Lynn & K. McConkey (Eds.), *Truth in memory* (pp. 109–136). New York: Guilford Press.

Marmelstein, L., & Lynn, S. J. (1999). Expectancies, group, and hypnotic influences on early autobiographical memory reports. *International Journal of Clinical and Experimental Hypnosis, 47,* 301–319.

Mazzoni, G. A., Loftus, E. F., & Kirsch, I. (2001a). Changing beliefs about implausible autobiographical memories. *Journal of Experimental Psychology: Applied, 7,* 51–59.

Mazzoni, G. A., Loftus, E. F., Seitz, A., & Lynn, S. J. (1999b). Creating a new childhood: Changing beliefs and memories through dream interpretation. *Applied Cognitive Psychology, 13,* 125–144.

Mazzoni, G. A., Lombardo, P., Malvagia, S., & Loftus, E. F. (1999).Dream interpretation and false beliefs. *Professional Psychology: Research and Practice 30*(1), 45–50.

Mazzoni, G., & Memon, A. (2003). Imagination can create false autobiographical memories. *Psychological Science, 14*(2), 186–188.

McHugh, P. R. (2008). *Try to remember: Psychiatry's clash over meaning, memory, and mind*. New York: Dana Press.

McConkey, K. M., & Jupp, J. J. (1986). A survey of opinions about hypnosis. *British Journal of Experimental and Clinical Hypnosis, 34,* 311–319.

McNally, R. J. (2005). *Remembering trauma*. Cambridge, MA: Harvard University Press.

McNally, R. J. (2007). Betrayal trauma theory: A critical appraisal. *Memory, 15,* 280–294.

McNally, R. J. (2012). Searching for repressed memory. In R. F. Belli (Ed.), *True and false recovered memories: Toward a reconciliation of the debate* (pp. 121–147). New York: Springer.

McNally, R. J., & Geraerts, E. (2009). A new solution to the recovered memory debate. *Perspectives of Psychological Science, 4*(2), 126–134.

McNally, R. J., Lasko, N. B., Claney, S. A., Macklin, M. L., Pitman, R. K., & Orr, S. P. (2004). Psychophysiological responding during script-driven imagery in people reporting abduction by space aliens. *Psychological Science, 15*(7), 493–497.

Meiselman, K. (1990). *Resolving the trauma of incest: Reintegraton therapy with survivors.* San Francisco: Jossey-Bass.

Memon, A., Meissner, C. A., & Fraser, J. (2010). The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16,* 340–372.

Meyersburg, C. A., Bogdan, R. G., Gallo, D. A., & McNally, R. J. (2009). False memory propensity in people reporting recollected memories of past lives. *Journal of American Psychology, 118*(2), 399–404.

Mills, A., & Lynn, S. J. (2000). Past-life experiences. In E. Cardena, S. J. Lynn, & S. Krippner (Eds.), *The varieties of anomalous experience* (pp. 283–314). Washington, DC: American Psychological Association.

Morris, E. K., Laney, C., Bernstein, D. M., & Loftus, E. F. (2006). Susceptibility to memory distortion: How do we decide it has occurred. *American Journal of Psychology, 119*(2), 255–274.

Mulhern, S. (1992). Ritual abuse: Defining a syndrome versus defending a belief. *Journal of Psychology and Theology, 20,* 230–232.

Nadel, L., & Jacobs, J. W. (1998). Traumatic memory is special. *Current Directions in Psychological Science, 7,* 154–157.

Nash, M. J. (1987). What, if anything, is regressed about hypnotic age regression?: A review of the empirical literature. *Psychological Bulletin, 102,* 42–52.

Nash, M. J., Drake, M., Wiley, R., Khalsa, S., & Lynn, S. J. (l986). The accuracy of recall of hypnotically age regressed subjects. *Journal of Abnormal Psychology, 95,* 298–300.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review, 111*(2), 486–511.

Neuschatz, J. S., Lynn, S. J., Benoit, G. E., & Fite, R. (2002). Hypnosis and memory illusions: An investigation using the Deese/Roediger and McDermott paradigm. *Imagination, Cognition and Personality*, *22*(1), 3–12.

Olio, K. A. (1994). Truth in memory. *American Psychologist, 49,* 442–443.

Olson, H. A. (1979). The hypnotic retrieval of early recollections. In H. A. Olson (Ed.), *Early recollections: Their use in diagnosis and psychotherapy* (pp. 163–171). Springfield, IL: Charles C Thomas.

Ost, J., Ghonouie, H., Cook, L., & Vrij, A. (2008). The effects of confederate influence and confidence on the accuracy of crime judgments. *Acta Psychologica, 128*(1), 25–32.

Ost, J., Wright, D. B., Easton, S., Hope, L., & French, C. C. (2013). Recovered memories, satanic abuse, dissociative identity disorder and dales memories in the UK: A survey of clinical psychologists and hypnotherapists. *Psychology, Crime and Law, 19*(1), 1–19.

Otgaar, H., Peters, M., & Howe, M. L. (2012). Dividing attention lowers children's but increases adults' false memories. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(1), 204.

Papanek, H. (1979). The use of early recollections in psychotherapy. In H. A. Olson (Ed.), *Early recollectons: Their use in diagnosis and psychotherapy* (pp. 223–229). Springfield, IL: Charles C Thomas.

Peterson, C. (2012). Children's autobiographical memories across the years: Forensic implications of childhood amnesia and eyewitness memory for stressful events. *Developmental Review, 32,* 287–306.

Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science, 8,* 437–441.

Pezdek, K., & Lam, S. (2007). What research paradigms have cognitive psychologists used to study "false memory," and what are the implications of the choices? *Consciousness and Cognition, 16*(1), 2–17.

Pigotti, M., & Thyer, B. (2011). Clinical social work and evidence-based practice: An introduction to the special issue. *Clinical Social Work Journal, 39*(4), 325–327.

Polusny, M. A., & Follette, V. M. (1996). Remembering childhood sexual abuse: A national survey of psychologists' clinical practices, beliefs, and personal experiences. *Professional Psychology: Research and Practice, 27,* 41–52.

Poole, D. A., Lindsay, D. S., Memon, A., & Bull, R. (1995). Psychotherapists' opinions, practices, and experiences with recovery of memories of incestuous abuse. *Journal of Consulting and Clinical Psychology, 68,* 426–437.

Pope, K. S. (1996). Memory, abuse, and science: Questioning claims about the false memory syndrome epidemic. *American Psychologist, 51,* 957–974.

Pope, K. S., & Brown, L. S. (1996). *Recovered memories of abuse: Assessment, therapy, forensics*. Washington, DC: American Psychological Association.

Porter, S., Birt, A. R., Yuille, J. C., & Lehman, D. R. (2000). Negotiating false

memories: Interviewer and rememberer characteristics related to memory distortion. *Psychological Science, 11*(6), 507–510.

Porter, S., Taylor, K., & ten Brinke, L. (2008). Memory for media: Investigation of false memories for negatively and positively charged public events. *Memory, 16,* 658–666.

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated childhood emotional events: Implications for the recovered memory debate. *Law and Human Behavior, 23,* 517–537.

Pyun, Y. D., & Kim, Y. J. (2009). Experimental production of past-life memories in hypnosis. *International Journal of Clinical and Experimental Hypnosis, 57,* 3, 269–278.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Roland, C. B. (1993). Exploring childhood memories with adult survivors of sexual abuse: Concrete reconstruction and visualization techniques. *Journal of Mental Health Counseling, 15,* 363–372.

Schooler, J. W., Ambadar, Z., & Bendiksen, M. (1997). A cognitive corroborative case study approach for investigating discovered memories of sexual abuse. In J. D. Read & D. S. Lindsay (Eds.), *Recollections of trauma: Scientific evidence and clinical practice* (pp. 379–388). New York: Plenum.

Scoboria, A., Lynn, S. J., Hessen, J., & Fisico, S. (2007). So that's why I don't remember: Normalizing forgetting of childhood events influences false autobiographical beliefs but not memories. *Memory, 15*(8), 801–813.

Seamon, J. G., Philbin, M. M., & Harrison, L. G. (2006). Do you remember proposing marriage to the Pepsi machine?: False recollections from a campus walk. *Psychonomic Bulletin and Review, 13*(5), 752–756.

Sharman, S. J., & Scoboria, A. (2009). Imagination equally influences false memories of high and low plausibility events. *Applied Cognitive Psychology*, *23*(6), 813–827.

Shobe, K. K., & Kihlstrom, J. F. (1997). Is traumatic memory special? *Current Directions in Psychological Science, 6,* 70–74.

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the U.S. population. *PLoS ONE, 6*(8), e22757.

Simons, D. J., & Chabris, C. F. (2012). Common (mis)beliefs about memory: A replication and comparison of telephone and Mechanical Turk survey methods. *PLoS ONE, 7*(12), e51876.

Sivec, H. J., Lynn, S. J., & Malinoski, P. T. (1997a). *Early memory reports as a function of hypnotic and nonhypnotic age regression.* Unpublished manuscript, State University of New York at Binghamton.

Sivec, H. J., Lynn, S. J., & Malinoski, P. T. (1997b). *Hypnosis in the cabbage patch: Age regression with verifiable events.* Unpublished manuscript, State University of New York at Binghamton.

Spanos, N. P. (1996). *Multiple identities and false memories: A sociocognitive perspective.* Washington, DC: American Psychological Association.

Spanos, N. P., Burgess, C. A., & Burgess, M. F. (1994). Past life identities, UFO abductions, and satanic ritual abuse: The social construction of "memories." *International Journal of Experimental and Clinical Hypnosis, 42,* 433–446.

Spanos, N. P., Burgess, C. A., Burgess, M. F., Samuels, C., & Blois, W. O. (1999). Creating false memories of infancy with hypnotic and nonhypnotic procedures. *Applied Cognitive Psychology, 13,* 201–218.

Spanos, N. P., Menary, E., Gabora, M. J., DuBreuil, S. C., & Dewhirst, B. (1991). Secondary identity enactments during hypnotic past-life regression: A sociocognitive perspective. *Journal of Personality and Social Psychology, 61,* 308–320.

Stafford, J., & Lynn, S. J. (2002). Cultural scripts, memory of childhood abuse and multiple identities: A study of role-play enactments. *Journal of Experimental and Clinical Hypothesis, 50*(1), 67–85.

Steblay, N. M., & Bothwell, R. K. (1994). Evidence for hypnotically refreshed testimony: The view from the laboratory. *Law and Human Behavior, 18,* 635–651.

Strange, D., Garry, M., Bernstein, D. M., & Lindsay, D. S. (2011). Photographs cause false memories for the news. *Acta Psychologica, 136,* 90–94.

Tavris, C. (1993, January 3). Beware the incest survivor machine. *New York Times Book Review*, pp. 1, 16–17.

van der Kolk, B. A. (1994). The body keeps the score: Memory and the evolving psychobiology of posttraumatic stress. *Harvard Review of Psychiatry, 1,* 253–265.

van der Kolk, B. A., Britz, R., Burr, W., Sherry, S., & Hartmann, E. (1994). Nightmares and trauma: A comparison of nightmares after combat with life-long nightmares in veterans. *American Journal of Psychiatry, 141,* 187–190.

van Husen, J. E. (1988). The development of fears, phobias, and restrictive patterns of adaptation follwing attempted abortions. *Pre- and Peri-Natal Psychology Journal, 2,* 179–185.

Wade, K. A., Garry, M., Nash, R. A., & Harper, D. N. (2010). Anchoring effects in the development of false childhood memories. *Psychonomic Bulletin and Review, 17*(1), 66–72.

Wade, K. A., Garry, M. A., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin and Review, 9*(3), 597–603.

Wade, K. A., Sharman, S. J., Garry, M., Memon, A., Mazzoni, G., Merckelbach, H., et al. (2007). False claims about false memory research. *Consciousness and Cognition*, *16*(1), 18–28.

Weiss, B. L. (1988). *Many lives, many masters*. New York: Simon & Schuster.

Woolger, R. J. (1988). *Other lives, other selves: A Jungian psychotherapist discovers past lives*. New York: Bantam Books.

Yapko, M. D. (1994). Suggestibility and repressed memories of abuse: A survey of psychotherapists' beliefs. *American Journal of Clinical Hypnosis, 36,* 163–171.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., et al. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory, 18*, 543–555.

CHAPTER NINE

Self-Help Therapy

Recent Developments in the Science and Business of Giving Psychology Away

Gerald M. Rosen, Russell E. Glasgow,
Timothy E. Moore, and Manuel Barrera Jr.

Recent developments in professionally developed self-help programs are best appreciated within a historical context. In its earliest form, "self-help" referred to the coming together of peers who would assist each other independently of professional assistance. Katz and Bender (1976) traced the beginnings of self-help peer efforts to 19th-century England. Over the next century, the phenomenon of peer self-help groups continued to expand (Jacobs & Goodman, 1989). Gartner and Riessman (1977) estimated that at least 500 self-help organizations were active in the United States in the 1970s, a figure that is now dwarfed by hundreds, if not thousands, of "chat" groups on the World Wide Web.

Self-help treatment books represent another form of guidance that was available to the public long before psychologists got involved. Ellis (1977) suggested that the oldest and best-selling self-help text was the Bible, a document developed without the assistance of mental health professionals. In more recent times, best-selling self-help books continue to be written by authors outside the health professions. Norman Vincent Peale's (1952) *The Power of Positive Thinking* was a best seller through much of the second half of the 20th century. Peale was a cleric, not a health professional. Among the best-selling nonfiction books in 2000 (Wall Street Journal, 2000) was *The Art of Happiness* by the Dalai Lama. That same year, an article in *Newsweek* (January 10, 2000) entitled "Self-Help U.S.A." observed: "Since Colonial times, Americans have devoured 'success literature,' those

pragmatic guides to a better life from authors including Ben Franklin, Dale Carnegie. . . . Today they're called self-help books, and they constitute a $563 million-a-year publishing juggernaut." Advice-giving by nonprofessionals continues to the present day. At the time this chapter was written, *The Wall Street Journal* (July 2012) listed among the top 10 best-selling nonfiction books *It Worked for Me,* a treatise by former U.S. Secretary of State Colin Powell.

Given the enormous popularity of self-help materials (e.g., see Norcross, Campbell, Grohol, Santrock, Selegea, et al., 2013) and their goal of helping people help themselves, it is not surprising that psychologists and other health care professionals have provided their share of advice. An early example of professional involvement is the text by a physician, Samuel Smiles (1881), entitled *Self-Help*. Dr. Smiles (1886) also wrote *Happy Homes and the Hearts That Make Them*, a delightful text that contained chapters on "The Art of Living," "Influence of Character," and "Helping One's Self." Edmund Jacobsen's (1934) You Must Relax is another self-help book well known to psychologists. Starker (1989) has provided a history of self-help books and influential authors, and Jack and Ronan's (2008) historical review also is of some interest.

In the 1970s, there was an explosive growth of do-it-yourself books. Book publications continued in the 1980s, now accompanied by self-help audiocassettes and videotape programs. A 1988 *New York Times* article reported that one company, Mind Communications Inc., sold more than 6 million dollars' worth of subliminal tapes in that year, a 10-fold increase in sales in just 2 years (Lofflin, 1988). During the 1970s and 1980s, the American Psychological Association also entered the business of developing, marketing, and promoting self-help audiocassettes, an issue discussed later in this chapter. In the 1990s, self-help instructional formats were expanded to include computer programs (Marks et al., 2003; Marks, Shaw, & Parkin, 1998; Newman, Consoli, & Taylor, 1997). Self-help over the Internet (e.g., Carlbring, Furmark, Steczko, Ekselius, & Andersson, 2006; Jerome & Zaylor, 2000; Klein & Richards, 2001; Strom, Pettersson, & Andersson, 2000, 2004) and smartphone applications (e.g., Abroms, Padmanabhan, Thaweethai, & Phillips, 2011; Breton, Fuemmeler, & Abroms, 2011) are the most recently developed venues for delivering self-administered treatments to the public, with this venue for providing treatment extremely likely to grow in the near future (Norcross, Pfund, & Prochaska, 2013).

The self-help industry has expanded its reach in other ways. For example, in the area of parenting skills, there used to be general books of advice by authors such as Benjamin Spock. By the 1980s, there were individualized audiotapes that parents could play to children before bedtime for the more specific purposes of eliminating fears and bed-wetting, or improving self-esteem. There was a book specifically targeted to help infants with colic (Ayllon & Freed, 1989), and another program was directed at issues involving toilet training (Azrin & Foxx, 1974). This trend toward greater

specificity of focus, coupled with multiple modalities for delivering instructional programs, helps to explain how the self-help movement has become such big business (Lofflin, 1988; Newsweek, 2000).

PSYCHOLOGY AND SELF-HELP IN THE 1970s

Although the history of self-help spans centuries, it was not until the 1970s that leading academic psychologists became involved to any serious extent in writing and promoting these programs. For example, Lewinsohn wrote on depression (Lewinsohn, Munoz, Zeiss, & Youngren, 1979), Mahoney and Brownell on weight loss (Mahoney & Mahoney, 1976; Brownell, 1980), Heiman and LoPiccolo on sexual dysfunction (Heiman, LoPiccolo, & LoPiccolo, 1976), Coates and Thoresen on insomnia (Coates & Thoresen, 1977), Lichtenstein on smoking cessation (Danaher & Lichtenstein, 1978), Zimbardo on shyness (Zimbardo, 1977), and Azrin on habit control (Azrin & Foxx, 1974; Azrin & Nunn, 1977). These individuals, and other prominent psychologists, contributed to what remains an unprecedented push by academicians to develop self-help therapies (Rosen, 1976a).

At first glance, the involvement of psychologists in the development of self-help materials would seem beneficial. Psychologists who provide advice to the public appear to be following George Miller's urgings to "give psychology away" (Miller, 1969, p. 1074). Miller had used this phrase in his 1969 Presidential Address to the American Psychological Association to clarify what he saw as the major social responsibility of his profession—*to learn how to help people help themselves*. Certainly, this was the spirit of self-help or "do-it-yourself" treatment books in the 1970s—a theme of social consciousness that fit the times.

In line with Miller's urgings, psychologists appeared to be in a unique position to contribute to the self-help movement. By virtue of their training, they were equipped to develop and evaluate self-help instructional programs. Systematic work in the area had the potential to make available tested self-help therapies that consumers could self-administer or therapists could employ as adjuncts to office-based interventions. No other professional group combined the skills and expertise that psychologists could bring to bear on the development of these programs. In the most utopian fantasy, psychology would bring a new dawn to the self-help movement, one in which empirically supported materials were available for specific targeted goals. Rosen (1977) observed that untested products have always been sold to unwary consumers. Rosen invited psychologists to imagine a group of professionals who adequately validated self-help books and educated consumers in their proper use. That same year, at an American Psychological Association symposium, Albert Ellis spoke to the great potential for improved human functioning that a set of scientifically researched and periodically revised do-it-yourself manuals could have (Ellis, 1977). This

enthusiasm permeated the 1970s—a time when psychologists rushed headlong into the self-help movement.

SELF-HELP DOESN'T NECESSARILY HELP

In addition to prominent psychologists developing self-help programs in the 1970s, a considerable amount of outcome research was conducted. Glasgow and Rosen (1978, 1982) located 117 studies or case reports from this time period that evaluated behaviorally oriented self-help instructional materials. This constituted a sizeable body of research for which psychologists were to be commended. Importantly, findings from these studies demonstrated that the task of "giving psychology away" was more complex and challenging than was initially thought.

One sobering finding that emerged from research in the 1970s was that techniques applied successfully by a therapist were not always self-administered successfully. For example, a study by Matson and Ollendick (1977) evaluated a book entitled, *Toilet Training in Less Than a Day* (Azrin & Foxx, 1974). The study found that four of five mothers in a therapist-administered condition successfully toilet trained their children, whereas only one of five mothers who used the book in a self-administered condition was successful. This study also revealed that unsuccessful self-administered interventions were associated with an increase in children's problem behaviors and negative emotional side effects between mothers and children. In other words, highly successful interventions based in a clinic or supervised by a therapist did not necessarily translate into a helpful do-it-yourself program. The implications of this finding are apparent: If 100,000 copies of *Toilet Training in Less Than a Day* were sold and Matson's and Ollendick's (1977) findings applied, then 20,000 children might be expected to benefit from the self-instructional program, an impressive result at extremely low cost. Unfortunately, this seemingly positive outcome would say nothing about the 80,000 parents who might be frustrated, if not angry, because their children were among the 80% who did not respond.

Matson and Ollendick's findings were not unique. Zeiss and Zeiss (1978) conducted a controlled outcome study on the treatment of premature ejaculation. Couples were assigned, on a random basis, to receive either self-administered treatment, minimal therapist contact, or therapist-directed treatment. As in earlier reports by Zeiss (1977) and Lowe and Mikulas (1975), treatment with only minimal therapist contact was effective. But of six couples who self-administered treatment in Zeiss's 1978 study, *none* successfully completed the program.

In yet another demonstration of the point that well-intentioned instructional materials are not necessarily effective, Rosen, Glasgow, and Barrera (1976) found that participants who were highly fearful of snakes, and able

to totally self-administer a written desensitization program, significantly reduced their anxiety reactions. This positive and encouraging outcome was tempered by the additional finding that 50% of subjects in the self-administered condition failed to carry out instructional assignments. Other studies on self-administered fear reduction had shown similar problems with adherence. For example, 14 of 29 eligible participants dropped out in Clark (1973); 5 of 11 dropped out in Marshall, Presse, and Andrews (1976); and two-thirds of subjects failed to complete their program in Phillips, Johnson, and Geyer (1972). Problems with low levels of follow-through or engagement first identified in the 1970s remain a significant problem, even in the newer interactive technologies and formats for delivering self-help instructions. For example, many of the newer interactive technologies have found a "law of attrition" (Eysenbach, 2005) that characterizes progressive dropoff in the percentage of users actively involved in web-based interventions over time.

Because the issue of follow-through was a major impediment to helping people help themselves, Barrera and Rosen (1977) attempted to increase treatment adherence. In this study, phobic subjects were randomly assigned to the original self-administered program used in the 1976 study or to a revised program with self-reward contracting. The addition of a self-reward contracting module to self-administered desensitization was consistent with self-management efforts promoted at the time (Mahoney & Thoresen, 1974). Yet, results from the study were a total surprise. As in the 1976 outcome study, 50% of participants completed the original program, and these participants substantially reduced their fears. However, in the revised program with self-contracting, the number of participants who followed the instructions dropped from 50% to 0%. In other words, no subject completed the new and "improved" program, resulting in an associated drop in treatment gains. The importance of this unanticipated finding cannot be overemphasized, for it clearly demonstrated that well-intentioned changes in instructional materials can have a significant and negative impact on adherence to procedures, and consequently on treatment outcome.

Findings from the studies cited above, together with additional research efforts in the 1970s, yielded several conclusions (Glasgow & Rosen, 1978):

1. Self-help instructional materials can be effective and provide a cost-effective alternative to therapist-directed interventions.
2. Self-help instructional materials may not be effective, even when based on empirically supported therapist-directed methods. In other words, the effectiveness of a treatment program under one set of conditions does not necessarily generalize to other conditions.
3. The effect of *any* change in instructional content, no matter how well intentioned, can produce unintended results and is an empirical question that must be assessed under the specific conditions for which materials are intended.

A RUSH TO PUBLISH

Despite research findings from the 1970s, psychologists continued to publish untested programs. The authors of *Toilet Training in Less than a Day* (Azrin & Foxx, 1974) even contracted with a manufacturer of musical toilet seats to produce a package of products entitled *Less Than a Day Toilet Trainer.* Azrin then published a new and untested book under the title *Habit Control in a Day* (Azrin & Nunn, 1977). Rosen revised his desensitization program yet another time and published *Don't Be Afraid* (Rosen, 1976b), despite findings from well-controlled studies showing follow-through rates as low as 0%. To appreciate within a historical context the 1976 publication of *Don't Be Afraid,* one only has to search that title to find an earlier *Don't Be Afraid* published by Edward Cowles in 1941. This older *Don't Be Afraid* differed in content from the 1976 publication in that it promoted nerve fatigue theories rather than "modern" desensitization. Yet, in the absence of supporting research, psychologists and consumers cannot know if any advance in the self-treatment of phobic disorders actually occurred with the passage of a quarter of a century. For all we know, the 1941 *Don't Be Afraid* is just as effective, or more effective, than any of the well-intentioned drafts Rosen developed in the 1970s.

In addition to rushing untested programs to market in an effort to "give psychology away," some psychologists allowed their programs to be accompanied by unsubstantiated claims. This observation may provide the most dramatic demonstration that commercial factors, rather than professional standards, often dominate the marketing of self-help books. Take, for example, the 1976 *Don't Be Afraid*, which stated on its book jacket stated: "In as little as six to eight weeks, without the expense of professional counseling, and in the privacy of your own home, you can learn to master those situations that now make you nervous or afraid" (Rosen 1976b). Note that research findings are not mentioned to clarify that, at best, 50% of people succeeded at self-administered treatment.

Or consider the commercial blurb on the back jacket of *Toilet Training in Less Than a Day* (Azrin & Foxx, 1974):

> Two noted learning specialists have developed this amazing new training method that—for the average child—requires less than four hours. . . . Parents report other benefits from the Azrin–Foxx scientifically tested method. . . . Many parents find they have more time for personal interests, and that their child is now a source of increased pride and pleasure.

As noted earlier, in their small sample, Matson and Ollendick (1977) found that self-administered training was successful only 20% of the time. Consequently, feelings of "pride and pleasure" may have been absent from 80% of the households who purchased *Toilet Training in Less Than a Day.*

Most claims made by publishers demonstrate a lack of constraint.

Consider claims provided on the back cover of *In the Mind's Eye* (Lazarus, 1977), a book that presented cognitive-behavioral strategies touted to help the reader heighten his/her "creative powers, stop smoking, drinking or overeating, overcome sadness and despondence, build self-confidence and skill, overcome fears and anxiety." Lazarus personally intervened and had the publisher drop these claims at the next printing of the text. But three years later, Jerome Singer, then Director of the Clinical Psychology program at Yale University, published *Mind Play: The Creative Uses of Fantasy* (Singer & Switzer, 1980), another book presenting cognitive-behavioral techniques. This time, according to the publisher's book jacket, a reader could "relax, overcome fears and bad habits, cope with pain, improve your decision-making and planning, perfect your skill at sports and enhance your sex life." Neither *In the Mind's Eye* nor *Mind Play* was ever assessed to determine if any of the claims were in reach of the typical consumer.

Flash forward to the present and little has changed. Consider, for example, *The Worry Cure: Seven Steps to Stop Worry from Stopping You* (2005). Robert Leahy, former president of the Association for Cognitive and Behavior Therapy, wrote this self-help instructional program. Consistent with the empirical origins of behavior therapy, the book largely presents methods that have been found effective in clinic settings. Yet the book itself has never been tested. Without any apparent concern for this lack of empirical support, the following claims are made on the book's back jacket:

> *The Worry Cure* is for everyone, from the chronic worrier to the occasional ruminator. It's time to stop the racing thoughts and start using the groundbreaking methods in this book to achieve the healthier, more successful life you deserve.

Does *The Worry Cure* fare any better than *Mind Power* circa 1903, or *Mind Play* circa 1980, when it comes to fulfilling promotional claims? Do 10%, 20%, or 50% of readers who self-administer *The Worry Cure* get better? Do they get worse? We simply don't know.

ENTHUSIASM FOR THE SELF-HELP INDUSTRY

If the 1970s represented a decade during which psychologists tried to "give psychology away," unencumbered by concerns over the therapeutic value of their gifts, then the following two decades represented a time when marketing strategies were refined, programs proliferated, and data remained sparse (Rosen, 1987, 1993). At the time of our first writing of this chapter (Rosen, Glasgow, & Moore, 2003), we found support for this appraisal by logging on to the web at *www.amazon.com*, where 137 self-help books were listed for just the letter A. Among the titles at *www.amazon.com* were *A.D.D. and Success, Access Your Brain's Joy Center:*

The Free Soul Method, Amazing Results of Positive Thinking, and *The Anxiety Cure: An Eight-Step Program for Getting Well*. There also were many titles with the word "Art" as in *The Art of Letting Go, The Art of Making Sex Sacred,* and *The Art of Midlife*. Findings were similar for the letters B through Z. A more recent visit to amazon.com demonstrated that titles for self-help books are every bit as ambitious in promoting their claims for success. Consider these two texts: *Achieve Anything in Just One Year* and *The Self-Help Bible: How to Change the Life You Have for the Life You Want*.

When writing the first edition of this chapter, we also visited PsycINFO, a search engine on the web that the American Psychological Association maintains to archive articles from major peer-reviewed journals. In response to the key word "Bibliotherapy," we found 60 records listed for the 1970s, 207 records for the 1980s, and 205 records for the 1990s. In a follow-up search to supplement our current efforts, we found 259 references on PsycINFO for the first decade of this century. Such findings suggest a continuing and active interest in self-help materials, with psychologists productively studying and advancing the development of these programs. Unfortunately, a more detailed inspection of the record is not encouraging. Take, for example, the bibliotherapy references for the time frame of 1990 through 1999. If one *excludes* from the 205 listed references all dissertations, chapters, commentaries, and review articles, and *includes* only controlled studies that actually assessed a self-help book, then the actual number of references for the entire decade dwindles to 15. This represents a paltry number of studies in relation to the thousands of volumes available from retailers. A similar lack of research characterizes the status of smartphone self-help applications for psychological problems (Breton et al., 2011).

The observation that research establishing evidence-based real-world applications of self-help programs has been limited mirrors a finding obtained many years ago by Glasgow and Rosen (1978, 1982). These authors conducted two reviews of the literature on behavioral self-help programs in the late 1970s and noted that the overall ratio of studies to books had dropped from 0.86 to 0.59, between the two reviews. This apparent decrease in research efforts was followed by a rise in publications that extolled the virtues of self-help (Ganzer, 1995; Johnson & Johnson, 1998; Lanza, 1996; Quackenbush, 1992; Warner, 1992). In fact, of the 205 references constituting the 1990s professional literature on bibliotherapy, there were more position papers urging psychologists to use these programs than controlled studies on their effectiveness. One author alone contributed 14 such references (Pardeck, 1990a, 1990b, 1990c, 1991a, 1991b, 1992a, 1992b, 1993, 1994, 1996, 1997; Pardeck & Pardeck, 1993, 1999; Pardeck & Markward, 1995). Perhaps the most enthusiastic example of this genre was a paper by Norcross (2000), "Here comes the self-help revolution in mental health," in which the author observed:

> The self-help revolution is here and it is growing. Psychologists can idly watch with bemused interest—devaluing self-change as shallow, self-help books as trivial, and Internet sites as harmless—as the train roars past us. Alternatively, we can recognize the power and potential of the locomotive and help steer it to valuable destinations for our patients and the populace. (pp. 375–376)

Enthusiasm for the use of self-help books has been supported, at least in part, by several meta-analytic studies that demonstrate the general effectiveness of tested programs (Gregory, Canning, Lee, & Wise, 2004; Erford et al., 2013; Gould & Clum, 1993; Kurtzweil, Scogin, & Rosen, 1996; Marrs, 1995; Scogin, Bynum, Stephens, & Calhoon, 1990). Moreover, some recent data regarding certain Internet-based self-help programs, such as those for patients with partially remitted depression, are potentially encouraging (Lintvedt et al., 2013). Nevertheless, these publications relate to a small subset of self-help programs only and add little to an overall appraisal of self-help interventions available in the marketplace. Further, the general conclusion that self-help books can be effective, but a particular program's value can only be established by testing it, has been known for some time (Glasgow & Rosen, 1978).

General reviews of self-help books in the form of consumer guides for the public came into existence in the 1990s. The *Authoritative Guide to Self-Help Resources in Mental Health* (Norcross et al., 2000) is the best known example of this genre. Such reviews are not based on actual outcome studies: instead, their recommendations are based on personal preferences and/or surveys that poll psychologists on the materials they like to use. Popularity polls among psychologists provide no useful information to the public regarding a consumer's ability to self-administer a program at home. A "one-to-five star" popularity rating system applied to an opinion survey falls far short of good science if the goal is to provide a sound basis on which consumers can choose an instructional program.

Redding, Herbert, Forman, and Gaudiano (2008) provided a review of 50 top-selling self-help books for anxiety, depression, and trauma-related disorders. Selected books were rated by "expert" psychologists with regard to perceived overall usefulness, extent to which instructions were grounded in psychological science, and extent to which explanations for treatment rationales and methods were given. In other words, ratings were provided absent any assessment of a program's actual clinical utility. Not surprisingly, raters drawn from a cognitive-behavioral orientation rated most highly those books that presented a similar perspective. One might arguably have obtained similar results from psychoanalytic, religious, or New Age therapists. The central problem with the Redding et al. study is that raters preferred programs whose content they viewed as "scientifically grounded," but the programs themselves were not scientifically grounded, if by that phrase we mean "empirically assessed." As we have known since

the 1970s, if a program is not actually assessed under the conditions for which it is intended, then we have no way of knowing how effective it is.

A FEW POSITIVE DEVELOPMENTS

Enthusiasm for self-help therapies notwithstanding, very little has changed over a span of four decades, at least for the consumer in need of demonstrated-to-be-effective instructional materials. Although thousands of programs have been marketed, the great majority remain untested. Even when we consider new modalities for delivering instructional programs (e.g., the Internet and cell phones), all the original issues that apply to self-help books remain (Glasgow, Boles, McKay, & Barrera, 2003). From an empirical perspective, the self-help field has largely stood still, even with the involvement of professionals over a span of four decades.

Still, several positive developments have taken place for which psychologists are to be credited. One area in which psychologists have clarified important issues through the conduct of systematic research concerns subliminal self-help programs. These programs started to appear in bookstores in the form of self-help audiotapes starting in the early 1980s. They shared a common format in that the only consciously perceivable sounds on the tapes consisted of music, ocean waves, and the occasional bird cry. The intended therapeutic effects were purportedly brought about by the unconscious (i.e., subliminal) perception of specific affirmations contained on the tapes. The range of problems that subliminal tapes could ostensibly alleviate was diverse and included weight loss, memory enhancement, breast enlargement, improvement of sexual function, and relief from constipation.

The notion of surreptitious mind control has always been a popular topic with journalists (Pratkanis, 1992). As Koshland (1991) noted, however, the ultimate criterion for resolving a scientific controversy must be the data in a well-run experiment. Without empirical support, claims of the sort made on behalf of subliminal self-help programs are no better and no different than similar declarations made on behalf of snake oil over 100 years ago (Young, 1961). The proposition that motives can be influenced in important ways by directives that are so faint that their mere presence (let alone their content) is undetectable is an extraordinary claim.

To the credit of psychologists, it did not take long for researchers to demonstrate that claims of subliminal therapeutic influence were unfounded (cf. Cleerman, Bayne, & Wilken, 2009). Subliminal perception is most appropriately defined as a situation in which there is a discrepancy between the viewer's phenomenal experience and his/her ability to discriminate between different stimuli. Thus, many laboratory studies have shown that people are often sensitive to stimuli they claim not to have perceived. When required to distinguish between two or more stimuli, subjects can do so with some success, even while professing to be guessing (Holender, 1986). Therefore, subliminal perception is not perception in the absence of

stimulus sensitivity. Rather, it occurs when subjective experience is at odds with objective measures of signal detection. Subliminal tapes represented a change in modality from visual to auditory. Claims about the utility of subliminal tapes are thus essentially claims about the subliminal perception of speech. It is not obvious what the analogue to visual masking is for a speech signal. Masking, in the visual domain, is procedurally defined with relative precision. The mask does not impair or change the target stimulus, it simply limits the time available for processing the preceding target. In the absence of the mask, the target is easily perceived. Research (Kouider & Dupoux, 2005) has shown that speech perception in the absence of conscious awareness is possible, although the nature of the effect does not extend to semantic activation. Semantic priming was achieved only when the prime stimuli were available to consciousness.

Because there is little reliable evidence of semantic processing of stimuli that cannot be discriminated (Cheesman & Merikle, 1986), a failure to demonstrate such discrimination would preclude any effects attributable to the semantic content of a word or message. With respect to subliminal tapes, Merikle (1988) showed that listeners were unable to distinguish a subliminal tape from a placebo control in a forced-choice task. This presence/absence discrimination required a "placebo" tape that was identical to its companion subliminal tape but without any subliminal message. Similarly, Moore (1995) used matched pairs of audiotapes from three different manufacturers and found that subjects could not discriminate between tapes containing ostensibly different subliminal messages. Merikle's and Moore's data are important because they strongly suggest that the subliminal content of the tapes tested does not trigger any perceptual activity.

Findings from studies on signal detection lead to the hypothesis that subliminal self-help "messages" should not lead to any therapeutic benefit because these programs do not contain a signal capable of triggering any perceptual activity—conscious or otherwise. The prediction of null treatment effects is exactly what controlled research has found. For example, in an innovative study by Pratkanis, Eskanazi, and Greenwald (1994) participants listened daily for five weeks to tapes designed to improve either self-esteem or memory. Unbeknownst to the subjects, half of them received mislabeled tapes. That is, half the subjects with self-esteem tapes actually listened to tapes designed to improve memory. Similarly, half the subjects who thought they had memory tapes were really listening to self-esteem tapes. Pre- and posttest measures of both self-esteem and memory revealed that use of the tapes brought about no improvements in either domain of functioning. Other investigators have similarly shown that subliminal messages do not contribute to improved study skills (Russell, Rowe, & Smouse, 1991), reductions in anxiety (Auday, Mellett, & Williams, 1991), or weight loss (Merikle & Skanes, 1992).

Semantic activation without conscious awareness is a well-established phenomenon, but proponents of subliminal tapes apparently jumped to the conclusion that subliminal directives could directly influence the systemic

unconscious. Unconscious perceptual processes were assumed to provide the means by which therapeutic directives could be smuggled into the unconscious through the back door. Scientific studies have made it clear that this assumption has neither theoretical nor empirical support.

In addition to sound research evaluating unfounded claims for subliminal self-help tapes, several systematic research programs provide support for instructional programs intended to be self-administered. A few examples are provided here, though the interested reader can refer to Watkins and Clum (2008) for more thorough reviews. Scogin and colleagues (Scogin, Jamison, & Gochneaur, 1989; Scogin, Jamison, & Davis, 1990) have shown that a book on depression (Burns, 1980) can assist older adults with mood problems. Psychologists interested in acceptance and commitment therapy (ACT) have conducted multiple studies assessing various self-help programs based on these methods (e.g., Jeffcoat & Hayes, 2012; Muto, Hayes, & Jeffcoat, 2011; Thorsell et al., 2011). Fairburn and colleagues (Fairburn, 1995), have investigated under a variety of conditions including "pure self-help," finding positive effects (Carter & Fairburn, 1998; Ghaderi & Scott, 2003; Mitchell et al., 2011; Striegel-Moore et al., 2010).

Clum and associates (Gould, Clum, & Shapiro, 1993; Gould & Clum, 1995; Lidran, Watkins, Gould, & Clum, 1995) assessed an instructional book that can be self-administered to manage panic disorder (Clum, 1990) and found support for efficacy in controlled studies. At the same time, a study by Febbararo, Clum, Roodman, and Wright (1999) found that a totally self-administered application of the program was *not* effective, thereby casting "doubt on the efficacy of bibliotherapy and self-monitoring interventions when utilized absent from contact with a professional who conducts the assessment and monitors treatment compliance" (p. 209). This finding is consistent with previously cited research from the 1970s (Mattson & Ollendick, 1977; Zeiss & Zeiss, 1978), in which effects associated with therapist-assisted programs did not generalize to self-administered conditions and findings from a meta-analysis (Marrs, 1995), in which amount of therapist contact was found to moderate outcome for individuals with anxiety problems. Significantly, these findings reinforce concerns arising out of research in the 1970s. Once again, researchers have demonstrated that the only way to determine the effectiveness of well-intentioned instructional materials is to test those materials in the specific context of their intended usage. Psychologists who write self-help materials based on methods they find effective in office or clinic settings have no assurance that the public can successfully apply these procedures on their own.

Recently, several computer-assisted therapies have received increased attention with promising findings (Kiluk et al., 2011; Marks et al., 2003; Marks, Shaw, & Parkin, 1998). Roy-Byrne, Craske, and colleagues demonstrated that the computerized delivery of evidence-based treatments for multiple anxiety disorders achieved better results in a primary care center than usual care (Roy-Byrne et al., 2010; Craske et al., 2011). Unlike self-help books, computer programs can present instructional modules that are

not only tailored to a particular concern, but also take into account a consumer's educational level and offer decision trees for progressing through a program. Continued efforts to promote real-world applications and alternative models of delivery (Gunter & Whittal, 2010) may help to "reboot" psychotherapy for the benefit of the public (Kazdin & Blase, 2011). Although we strongly support further investigations that advance new models for delivering services, we also urge caution when drawing conclusions. In this context, researchers need to ask infrequently addressed questions such as: What percent and what types of consumers will purchase and use a self-help program, and what outcomes for what types of patients are produced under what conditions? (Bennett & Glasgow, 2009; Pawson, Greenhalgh, Harvey, & Walshe, 2005).

PROFESSIONAL ORGANIZATIONS AND THEIR FAILURE TO LEAD

In spite of a few positive developments arising from systematic research efforts, and a better understanding of the potential benefits and limitations of self-help instructional materials, several aspects of the self-help marketplace have not improved significantly over the years. For instance, research findings have not led to the demise of subliminal audiocassettes. At the time this chapter was written, amazon.com had 405 results for subliminal tapes with such titles as "Slim Forever," "Sleep Soundly," "Effortless Relaxation," "Build Self-Esteem," and "Quit Smoking." A visit to the "Self-Help" section of any local bookstore convincingly demonstrates that untested books of advice also flourish. In the midst of this ever-continuing market expansion, psychologists have added to the glut of untested programs more than they have advanced the empirical foundations of self-help.

When Miller (1969), more than 40 years ago, urged psychologists to "give psychology away," his admonition was to promote "human welfare" and encourage the systematic development and assessment of effective self-help methods. Miller was not encouraging the headlong rush to market of untested materials that has characterized the behavior of most authors over a 40-year period. In one sense, of course, there is nothing wrong with selling programs of advice. Certainly, everyone has the right to market whatever wisdom or guidance they wish to dispense. At the same time, psychologists who publish untested programs with misleading titles and unwarranted claims are not meeting high professional standards, nor are these individuals applying the science of psychology for the advancement of self-care.

Psychologists who use the status of their profession to promote untested self-help programs invite the public to be skeptical of science (Rosen, 1987, 1993). Robitscher (1980) expressed this concern while addressing a psychiatric audience:

> Every commercial exploitation of psychiatry, large or small, detracts from an integrity that psychiatry needs if it is to have meaning . . .

> when it becomes commercial, psychiatry dwindles down to a treatment of symptoms and an exploitation of techniques, a pretense of helping another that helps only the self. Many psychiatrists do not approve of the commercialism of psychiatry . . . but almost no psychiatrist speaks out against it. They turn their eyes away to avoid the sight of the money tree being shaken. . . . In the absence of psychiatrists who do not exploit psychiatry, those who do flourish.

There is little indication that the present situation is changing. In the 1970s, interested groups within the American Psychological Association formed Task Forces on Self-Help Therapies. The Task Forces issued recommendations in 1978 and 1990 that suggested the American Psychological Association take the following actions:

1. Develop a set of guidelines for psychologists similar to the standards that guide developers of psychological test materials. Such guidelines could clarify methodological and outcome evaluation issues pertinent to the adequate development of self-help therapies.
2. Provide psychologists with a list of informational points that should be included in a commercially available self-help program. For example, books would contain a front page that discussed the extent to which the program was evaluated, recommended uses of the program. and established reading level of the written instructions.
3. Provide a set of guidelines to aid psychologists who negotiate with publishers. The publication of sample contract clauses could significantly improve the position of psychologists who wish to set limits on claims or other promotional efforts.
4. Develop a short pamphlet to educate the public in the use of self-help therapies. The public could be informed as to how self-help therapies are used as adjuncts to therapist-assisted treatment or by themselves. The issue of developing realistic expectancies in light of sensationalized claims could be addressed.
5. Consider working in concert with other professional or consumer-advocate groups in an effort to educate the consumer public and possibly develop a review process to review current evidence on self-help programs. In time, it was suggested, standards for establishing a formal "approval seal" might be possible.

The sponsoring groups who originated the Task Forces on Self-Help Therapies did not endorse any of these listed recommendations (Rosen, 1993, 1994). Perhaps more significantly, the membership of the American Psychological Association was itself—perhaps unwittingly—involved in the development, marketing, and promotion of untested self-help materials. This situation came about through the American Psychological Association's 1983 purchase of *Psychology Today* and the companion *Psychology Today* Tape Series. By 1985, psychologists on the staff of *Psychology*

Today were contracting for new audiotapes to be added to the series. A consumer could order Personal Impact, in which "clinical psychologist Cooper helps listeners become aware of and enhance their self-presentation to improve the impact they make on others." Under the section "Becoming More Self-Reliant," the potential consumer was told, "You (can) become a more attractive, appealing person." Under Mental Imagery, developed by Lazarus, the consumer was told: "Harness the powers of your mind! A noted psychologist explains how to use mental imagery to increase self-confidence, develop more energy and stamina, improve performance and proficiency, cope more effectively, overcome fears, and lose weight." The consumer who ordered one of these untested tapes also received a brochure with the name of the American Psychological Association on the front cover. On the back of this brochure, it stated, "backed by the expert resources of the 87,000 members of the American Psychological Association, the *Psychology Today* Tape Program provides a vital link between psychology and you."

By 1988, the American Psychological Association Board of Directors had disengaged from *Psychology Today* and sold the magazine to another publisher. Thus, for at least 3 years, the most prominent professional organization representing psychologists actively sought, produced, and promoted untested self-help materials accompanied by unsubstantiated claims that were purportedly backed (without membership approval) by the then 87,000 members. By participating in these activities, the American Psychological Association not only turned its eyes away from the "money tree" noted by Robitscher, but, for a period of time, the American Psychological Association was itself harvesting the tree's fruits. Further, by developing and marketing untested self-help tapes, the American Psychological Association failed to provide a model or higher standard for its members, some of whom were publishing their own untested programs.

In more recent years, starting in 2009, the Association for Behavioral and Cognitive Therapies (ABCT; formerly the Association for Advancement of Behavior Therapy), has awarded "self-help seals of merit" to selected self-help books. A visit to ABCT's website, at the time this chapter was written, found that 96 self-help books had received the organization's *Seal of Merit*. Although some of the listed books have been assessed in controlled outcome studies, the vast majority have not received empirical support. How could it be that the leading professional organization for the advancement of behavioral and cognitive therapies has been giving seals of merit to untested programs? The answer to this question is found by looking at the criteria ABCT uses to hand out their honorary citations (*www.abct.org/SHBooks/?shInfo=Pro;*):

> To earn the ABCT Self-Help Book Seal of Merit, a book must meet the following criteria:
>
> - employ cognitive and/or behavioral principles
> - discuss cognitive and/or behavioral principles or theories explicitly in text

- have documented empirical support that lends support for the methods presented
- include no suggestions or methods that are contraindicated by scientific evidence
- present treatment methods that have consistent evidence for their effectiveness (books describing methods without a consistent track record of empirical support, or mixed evidence, would not be eligible)
- be consistent with best practices

To their credit, ABCT's criteria attempt to bring empirically supported methods to the public. Sorge, Moore, and Toplak (2009) have illustrated methods for assessing self-help programs along these lines. Nevertheless, nowhere in ABCT's criteria is there any mention that a book must be effective under the conditions that apply to a real-life consumer. Under ABCT's guidelines, Azrin and Foxx's *Toilet Training in Less Than a Day,* which was empirically demonstrated to fail to help 80% of its readers in a self-administered condition, could qualify for ABCT's merit seal. So could Rosen's *Don't Be Afraid,* which in one of its earlier iterations had been shown to help no one. In effect, what ABCT has done is to craft yet another subjective rating system for approving the contents of a book. As was the case with the survey by Redding et al. (2008), the real meaning of "scientifically grounded," in terms of the actual effectiveness of a self-help program, has been lost.

THE LIMITS OF GOOD INTENTIONS

The core lesson learned from early research on self-help therapies is that good intentions do not necessarily equate with effective therapies. This central point continues to be demonstrated. Ehlers and colleagues randomly assigned motor vehicle accident survivors diagnosed with posttraumatic stress disorder to receive therapist-administered cognitive therapy, a self-help booklet, or repeated assessments (Ehlers et al., 2003). Although cognitive therapy was demonstrated to be effective, the self-help booklet was not superior to simple repeated assessments. Of greater concern, outcome for the self-help group was actually worse on two measures. The important caveat that sound clinical practice cannot be based on good intentions alone has been demonstrated in other areas of clinical practice (Lilienfeld, 2007). Drawing from research findings on critical incident stress debriefings (CISD), Devilly and Cotton (2004) reached a conclusion that applies just as well to self-help: "It is becoming clear that 'belief' may indeed be a dangerous emotion when coming to judge the effectiveness of an intervention" (p. 37).

Research on CISD also highlights the importance of measuring the right variable when a method is said to be "scientifically grounded." This is because consumers polled after debriefings report satisfaction with the methods. This

is very much akin to psychologists who rate how much they like a particular self-help program. Unfortunately, liking something doesn't equate with positive outcomes. The important variable on which interventions must be scientifically grounded is clinical efficacy in real-world applications.

The issues we are discussing in relation to self-help and the limits of good intentions are broadly based. Glasgow, Lichtenstein, and Marcus (2003) observed that many interventions for health problems that are efficacious in traditional efficacy-based studies do not translate into effective programs when delivered under real-world conditions. They noted several reasons for this finding and stressed that medical, behavioral, and public health interventions need to be tested with representative samples, in representative settings, under representative conditions of use.

THE FUTURE OF SELF-HELP

In looking to the future, it appears that early recommendations to advance psychology's contributions to self-help require modification. These recommendations focused on programs that were likely to be developed by individual psychologists who worked in a specialized area of clinical expertise. The general notion was that the individual psychologist would assume responsibility for the proper development and assessment of self-help instructional materials, and that professional organizations such as the American Psychological Association would assist psychologists by providing guidelines for negotiating with publishers and assist the consumer by providing guidelines for how best to choose among available programs. This model for promoting empirically supported self-help materials has failed over the past four decades. An alternative model is needed.

In contrast to an "individualistic" approach to the development and evaluation of self-help materials, a "public health" approach is more likely to advance the efficacy of these programs. Such an approach would employ three of the key characteristics of public health: (1) "transdisciplinarity;" (2) an emphasis on the reach and breadth of treatment effects; and (3) attention to the social-environmental context (Abrams et al., 1996; Brownson, Remington, & Davis, 1998; Winett, King, & Altman, 1989). The first of these characteristics, "transdisciplinarity," involves a team of professionals from diverse professions who collaborate to develop a program. Transdisciplinary approaches to self-help are needed because multiple factors, in addition to program content, influence the availability, use, and results of these materials. These factors include marketing considerations, the framing of health messages, literacy and readability, and the family and sociomedical context in which a program is used. Consequently, there are greater opportunities for contributions from professionals in health communications, marketing, multicultural studies, and other health professions. Consider, for example, the topic of weight loss, one of the single most popular self-help

topics. We have learned over the past decades that eating behavior and metabolic outcomes have numerous genetic, physiologic, nutritional, exercise, and social antecedents in addition to the core psychological and behavioral processes addressed by psychologists. Research programs that have continuity and address these issues within a broad multidisciplinary perspective may be more likely to systematically advance the development of an empirically based self-help weight-loss program, as compared to individual "leading-figure" psychologists who write their well-intentioned but untested books of advice, only to be replaced by the next and most current "authority."

Instead of placing the possibly unreasonable burden on a single author for evaluation of a self-help program, the empirical basis for effective self-help programs will be advanced more rapidly by having programs tested by a variety of individuals, in a variety of settings, and under a variety of conditions. For example, if a national group of educators, family physicians, or researchers were to decide that a given health topic was appropriate for self-help intervention, then members of related professional organizations, clinics, health maintenance organizations, or health care systems could coordinate multiple-site studies and pool their data. Examples of such multidisciplinary collaboration are available from the interactions among multiple scientists, including several psychologists, in formulation of the evidence-based guidelines on smoking cessation (Fiore, Jorenby, & Baker, 1997) and development of implementation guidelines by the Agency for Health Care Quality Research (*www.ahqr.gov*).

A second key feature of a public health approach is focus on the breadth and reach of an intervention program (Glasgow, Vogt, & Boles, 1999; Oldenburg, Hardcastle, & Kok, 1997). This perspective focuses on the consumer and can be contrasted with current self-help programs that generally have been developed without thorough consumer input. Self-help programs are more likely to attract and maintain the involvement of users to the extent that the program addresses the concerns and needs of a given group of consumers and can present information and strategies in a way that makes sense from their worldview, personal model, or illness representation (Hampson, 1996; Leventhal & Diefenbach, 1991). In particular, there is concern regarding whether an intervention reaches those most in need or only the relatively healthy, affluent individuals who have sufficient time and resources to devote to a program (Conrad, 1987; Glasgow, Eakin, & Toobert, 1996). This concern translates both into suggestions for design and distribution of self-help books, as well as evaluation criteria. Glasgow, Vogt, and Boles (1999) suggested that health promotion researchers need to "RE-AIM" their evaluations to explicitly consider the issues of Reach, Efficacy, Adoption (across different settings and professionals), Implementation, and Maintenance of intervention effects. These criteria apply equally well to self-help psychology programs.

The third important characteristic of a public health approach is attention to social-environmental context. As applied to self-help programs,

social context issues include whether instructional materials are used as a stand-alone intervention or supplemented by therapist or peer contact. We noted previously that some self-help books were found to be effective when used with therapist support, but not when used alone. It may also be that a book given to patients by their physician or therapist, whom patients understand will check on their progress, may be more effective than one they pick up at a bookstore. Other contextual factors include adjunctive therapeutic modalities such as proactive or reactive telephone support (e.g., Lichtenstein, Glasgow, Lando, Ossip-Klein, & Boles, 1996), the use of computer technology or "expert systems" to personalize or tailor intervention (Abrams, Mills, & Bulger, 1999), and an ever increasing array of other modalities such as the Internet, videotape or CD-ROM materials, and World Wide Web chat rooms (Norcross et al., 2013). Specification of the conditions under which a self-help program is effective and not effective will advance the development of empirically based self-help approaches and lead to development of a more sophisticated "stepped care, matched intervention approach" (Abrams et al., 1996; Brownell & Wadden, 1992) in which an initial assessment recommends conditions of administration likely to be most cost-effective for a given individual.

We acknowledge that a team and organizational approach to developing self-help programs may be fraught with challenges. All working models are subject to financial interests, departures from the methods and values of sound science, and breaches of professional standards. Nevertheless, we firmly believe that the real future of empirically supported instructional programs is to be found with program-based methods.

GUIDELINES FOR PSYCHOLOGISTS AND CONSUMERS

Authors of good will, religious leaders, and health professionals will continue to write books just as they always have done. Nowadays, well-intentioned authors can expand their advice-giving efforts through the use of audio- and videotapes, computerized programs, the Internet, and phone apps. Publishers will continue to promote these instructional materials, as they have done for many years, often accompanying their products with unwarranted titles and claims. Of course, "business as usual" in the self-help industry does not ensure that this year's instructional program will be more effective than the content of last year's advice.

In the 1970s, there was a sense of great optimism that the science of psychology was in a unique position to contribute to the advancement of self-help therapies. Recommendations were made to encourage psychologists to use their unique research and clinical skills to develop and promote empirically supported self-help programs. With the wisdom of hindsight and 40 years of experience, we now see that earlier recommendations made to psychologists who wanted to "give psychology away" were overly

optimistic. The notion that individual psychologists would carry the burden of assessing and improving their programs, while a professional organization such as the American Psychological Association would assist with supportive guidelines, has not been realized.

Self-help has not advanced substantially over the past four decades, and it is unlikely to advance over the next 40 years if prevailing models are maintained. Unless a new direction is taken, there is no reason to expect that the next *Don't Be Afraid*, published perhaps in the year 2025, will be any more effective than the *Don't Be Afraid*'s of 1976 and 1941. It is in this context that we provide guidelines for psychologists interested in advancing the empirical status of self-help therapies, and recommend a new, broader, and more inclusive approach to the development, use, and evaluation of self-help therapies. Rather than focusing all the responsibility on an individual author of a self-help program, a public health approach to self-help is strongly encouraged. This broader-based approach involves the coordinated efforts of health organizations, clinician groups, government agencies, and professional societies. Based on this approach and the consideration of "who benefits under what conditions," we have developed a checklist of questions (see Table 9.1) to help developers of self-help programs address key issues before marketing their programs. This table uses the RE-AIM framework previously discussed to organize questions under the headings of Reach, Effectiveness, Adoption, Implementation, and Maintenance (Gaglio & Glasgow, 2012; *www.re-aim.org*).

The checklist provided in Table 9.1 can also help consumers who are considering adoption of a particular program and want to consider the full range of issues that may affect their selection. However, since the vast majority of current self-help products remain untested, a consumer interested in self-change must follow a few very simple rules. First, the consumer can take comfort in the notion that most self-help products are inexpensive, and in that regard, there is little financial harm when buying the product. Second, the consumer should appreciate the fact that claims made for the product are not to be taken seriously unless independent empirical evidence supports the claims. This point is true even when the author of a program is a noted authority within a professional group, such as psychology or psychiatry. Third, the consumer should not feel badly or experience any self-blame if the instructional materials are difficult to apply, or not helpful when applied. Like the 80% of mothers who could not use on their own a toilet training procedure for their children (Matson & Ollendick, 1977), the 100% of males who could not successfully self-administer a program for sexual dysfunction (Zeiss, 1978), the 100% of snake phobics who failed to implement a self-administered desensitization procedure with self-reward contracting (Barrera & Rosen, 1977), and the panic disorder patients who failed to benefit from their self-administered program (Febbraro et al., 1999), the consumer may be dealing with an untested product that simply is not written in a manner that people can use.

TABLE 9.1. Guidelines for Developing, Selecting, or Evaluating a Self-Help Program: Questions to Ask

Reach (How broadly applicable is the program?)

- What percentage of the population having the particular problem, goal, or diagnosis is this program designed to address?
- Are there subgroups that are more or less likely to participate in this type of program?
- Are there data on the percentage of individuals who were offered this program who tried it? If yes:
  - What percent participated, and were they different from those who declined?

Efficacy (How effective is the program?)

- Has this program been evaluated? If yes:
  - Did it do better than a randomized or other type of control condition?
  - Did the program produce improvements on objective measures of outcome?
  - Were results reported for all persons who began the program—or only those who liked it and finished?
- Has the program been evaluated for possible negative or unwanted side effects? If so, what were these?
- Under what conditions has the program been administered? (Do *not* assume that results will be the same under different conditions.)
  - Completely self-administered:
  - Minimal therapist contact
  - As a supplement to regular counseling
- Under what modalities has the program been tested (e.g., written form; audio- or videotape; computer administered)?
- What is the cost of the program—both for purchase and amount of time required relative to alternative programs?
- Does the evidence for the program appear to match the claims that are made of it?

Adoption (How broadly has the program been used by groups other than the authors—and have the results of these other groups been equally positive?)

- Is there any information on the range of groups of clinicians, health systems, or researchers who have used or tested the program?
- Is there any information on the types of professionals or organizations that are likely to use versus not use this program?

Implementation (How easy is it to use the program?)

- What percentage of the initial users of this program complete the program, and how are they different from those who do not?

(*continued*)

TABLE 9.1. (*continued*)

- Are there any patient, setting, or procedural considerations for which this program seems to work best?
- Is there any way to get consultation or technical assistance with the program, if needed?

Maintenance (Does the program produce long-term or lasting results?)

- What are the longest follow-up assessments that have been conducted, and does the program still seem effective at longer-term follow-ups?
- Have the organizations or clinicians that have used the program continued to use it?

In closing, we want to recall the hopes for self-help that were expressed in the 1970s and referenced earlier in this chapter. In 1977, Ellis and Rosen invited psychologists to imagine the great potential a set of scientifically based do-it-yourself manuals could have. Many other psychologists shared the view that psychologists could meaningfully contribute to the development of effective and empirically based self-help programs, thereby fulfilling Miller's (1969) directive to promote human welfare by "giving psychology away." More than four decades later, we continue to support the idealism of the 1970s, and we continue to believe that psychologists will play an important role in the development of effective self-help materials. Imagine, if you will, a multidisciplinary group of professionals who develop self-help programs, educate consumers in their proper use, and continually evaluate and improve these programs in the context of long-term public health projects. The future of an empirically sound self-help movement lies in this vision of "program-based" methods rather than "individually authored" products.

GLOSSARY:

Adoption: The percentage and representativeness of professionals (or medical groups, clinics, health systems, and so on) who will use a given intervention or self-help program.

Bibliotherapy: The use of written materials (e.g., books, manuals) to further a personal goal or therapeutic objective.

Breadth: The range of applicability of a program; in this case, how broad a cross section of patients and providers will use and benefit from the program.

Compliance: The extent to which a patient follows professional advice. This term has largely been superseded by alternatives such as "self-management," which suggests a more central role for the patient in behavior change (Glasgow & Anderson, 1999).

Program completion: The percentage and representativeness of persons beginning a program who complete the intervention and follow its recommendations. This term, like "self-management," is preferred to the term "compliance."

Reach: The percentage of persons with a given condition or problem who try a given

approach or intervention, and the representativeness of this group of the entire population exhibiting this problem.

Self-help: The efforts of an individual to achieve behavior change or other personal goals without professional assistance.

Social-environmental context: The setting in which persons live (their family, neighborhood, cultural group, income level) and in which a program is used (e.g., purchased at a bookstore, used as part of therapy with a professional).

Subliminal: Commonly thought of as referring to the presentation of a stimulus below a threshold of conscious awareness, this term is better defined as a discrepancy between viewers' phenomenal experience and their ability to discriminate among different stimulus states.

REFERENCES

Abrams, D. B., Mills, S., & Bulger, D. (1999). Challenges and future directions for tailored communications research. *Annals of Behavioral Medicine, 21,* 299–306.

Abrams, D. B., Orleans, C. T., Niaura, R. S., Goldstein, M. G., Prochaska, J. O., & Velicer, W. (1996). Integrating individual and public health perspectives for treatment of tobacco dependence under managed care: A combined stepped care and matching model. *Annals of Behavioral Medicine, 18,* 290–304.

Abroms, L. C., Padmanabhan, N., Thaweethai, L., & Phillips, T. (2011). iPhone apps for smoking cessation: A content analysis. *American Journal of Preventive Medicine, 40,* 279–285.

Auday, B. C., Mellett, J. L., & Williams, P. M. (1991, April). *Self-improvement using subliminal self-help audiotapes: Consumer benefit or consumer fraud?* Paper presented at the annual meeting of the Western Psychological Association, San Francisco.

Ayllon, T., & Freed, M. (1989). *Stopping baby's colic*. New York: Putnam.

Azrin, N. H., & Foxx, R. M. (1974). *Toilet training in less than a day*. New York: Simon & Schuster.

Azrin, N. H., & Nunn, R. G. (1977). *Habit control in a day*. New York: Simon & Schuster.

Barrera, M., Jr., & Rosen, G. M. (1977). Detrimental effects of a self-reward contracting program on subjects' involvement in self-administered desensitization. *Journal of Consulting and Clinical Psychology*, *45*, 1180–1181.

Bennett, G. G., & Glasgow, R. E. (2009). The delivery of public health interventions via the Internet: Actualizing their potential. *Annual Review of Public Health, 30,* 273–292.

Breton, E. R., Fuemmeler, B. F., & Abroms, L. C. (2011). Weight loss—there is an app for that!: But does it adhere to evidence-informed practices? *Translational Behavioral Medicine, 1,* 523–529.

Brownell, K. D. (1980). *The partnership diet program*. New York: Rawon.

Brownell, K. D., & Wadden, T. A. (1992). Etiology and treatment of obesity: Understanding a serious, prevalent, and refractory disorder. *Journal of Consulting and Clinical Psychology, 60,* 505–517.

Brownson, R. C., Remington, P. L., & Davis, J. R. (1998). *Chronic disease epidemiology and control* (2nd ed.). Washington, DC: American Public Health Association.

Burns, D. (1980). *Feeling good*. New York: Avon Books.
Carlbring, P., Furmark, T., Steczko, J., Ekselius, L., & Andersson, G. (2006). An open study of Internet-based bibliotherapy with minimal therapist contact via e-mail for social phobia. *Clinical Psychologist, 10,* 30–38.
Carter, J. C., & Fairburn, C. G. (1998). Cognitive-behavioral self-help for binge eating disorder: A controlled effectiveness study. *Journal of Consulting and Clinical Psychology, 66,* 616–623.
Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology, 40,* 343–367.
Clark, F. (1973). Self-administered desensitization. *Behaviour Research and Therapy, 11,* 335–338.
Cleermans, A, Bayne, T., & Wilken, P. (Eds.). (2009). Subliminal tapes. In *The Oxford companion to consciousness*. Oxford, UK: Oxford University Press.
Clum, G. A. (1990). *Coping with panic*. Pacific Grove, CA: Brooks/Cole.
Coates, T., & Thoresen, C. E. (1977). *How to sleep better: A drug-free program for overcoming insomnia*. Englewood Cliffs, NJ: Prentice-Hall.
Conrad, P. (1987). Who comes to worksite wellness programs?: A preliminary review. *Journal of Occupational Medicine, 29,* 317–320.
Cowles, E. S. (1941). *Don't be afraid!* New York: McGraw-Hill.
Craske, M. G., Stein, M. B., Sullivan, G., Sherbourne, C., Bystritsky, A., et al. (2011). Disorder-specific impact of coordinated anxiety learning and management treatment for anxiety disorders in primary care. *Archives of General Psychiatry, 68,* 378–388.
Danaher, B. G., & Liechtenstein, E. (1978). *Become an ex-smoker.* Englewood Cliffs, NJ: Prentice-Hall.
Devilly, G. J., & Cotton, P. (2004). Caveat emptor, caveat venditor, and critical incident stress debriefing/management (CISD/M). *Australian Psychologist, 39,* 35–40.
Ehlers, A., Clark, D. M., Hackmann, A., McManus, F., Fennell, M., Herbert, C., et al. (2003). A randomized controlled trial of cognitive therapy, a self-help booklet, and repeated assessments as early interventions for posttraumatic stress disorder. *Archives of General Psychiatry, 60*, 1024–1032.
Ellis, A. (1977, August). Rational-emotive therapy and self-help therapy. In G. M. Rosen (Chair), *Non-prescription psychotherapies: A symposium on do-it-yourself treatments*. Symposium conducted at the meeting of the American Psychological Association, San Francisco, CA.
Erford, B. T., Richards, T., Peacock, E., Voith, K., McGair, H., Muller, B., et al. (2013). Counseling and guided self-help outcomes for clients with bulimia nervosa: A meta-analysis of clinical trials from 1980 to 2010. *Journal of Counseling and Development, 91,* 152–172.
Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7,* e11.
Fairburn, C. (1995). *Overcoming binge eating.* New York: Guilford Press.
Febbraro, G. A. R., Clum, G. A., Roodman, A. A., & Wright, J. H. (1999). The limits of bibliotherapy: A study of the differential effectiveness of self-administered interventions in individuals with panic attacks. *Behavior Therapy, 30,* 209–222.
Fiore, M. C., Jorenby, D. E., & Baker, T. B. (1997). Smoking cessation: Principles and practice based upon the AHCPR guidelines, 1996. *Annals of Behavioral Medicine, 19,* 213–219.

Gaglio, B., & Glasgow, R. E. (2012). Evaluation approaches for dissemination and implementation research. In R. Brownson, G. Colditz, & E. Proctor (Eds.), *Dissemination and implementation research in health* (pp. 327–356). New York: Oxford University Press.

Ganzer, C. (1995). Using literature as an aid to practice. *Families in Society, 75,* 616–623.

Gartner, A., & Riessman, F. (1977). *Self-help in the human services*. San Francisco: Jossey-Bass.

Ghaderi, A., & Scott, B. (2003). Pure and guided self-help for full and sub-threshold bulimia nervosa and binge eating disorder. *British Journal of Clinical Psychology, 42,* 257–269.

Glasgow, R. E., & Anderson, R. M. (1999). In diabetes care, moving from compliance to adherence is not enough: Something entirely different is needed. *Diabetes Care, 12,* 2090–2092.

Glasgow, R. E., Boles, S., McKay, G., Feil, E., & Barrera, M. (2003). The D-Net diabetes self-management program: Long-term implementation, outcomes, and generalization results. *Preventive Medicine*, *36*, 410–419.

Glasgow, R. E., Eakin, E. G., & Toobert, D. J. (1996). How generalizable are the results of diabetes self-management research?: The impact of participation and attrition. *Diabetes Educator, 22,* 573–585.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice?: Rethinking the efficacy to effectiveness transition. *American Journal of Public Health, 93,* 1261–1267.

Glasgow, R. E., & Rosen, G. M. (1978). Behavioral bibliotherapy: A review of self-help behavior therapy manuals. *Psychological Bulletin, 85,* 1–23.

Glasgow, R. E., & Rosen, G. M. (1982). Self-help behavior therapy manuals: Recent development and clinical usage. *Clinical Behavior Therapy Review, 1,* 1–20.

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health, 89,* 1322–1327.

Glasgow, R. E., Whitlock, E. P., Eakin, E. G., & Lichtenstein, E. (2000). A brief smoking cessation intervention for women in low-income planned parenthood clinics. *American Journal of Public Health, 90,* 786–789.

Gould, R. A., & Clum, G. A. (1993). A meta-analysis of self-help treatment approaches. *Clinical Psychology Review, 13,* 169–186.

Gould, R. A., & Clum, G. A. (1995). Self-help plus minimal therapist contact in the treatment of panic disorder: A replication and extension. *Behavior Therapy, 26,* 533–546.

Gould, R. A., Clum, G. A., & Shapiro, D. (1993). The use of bibliotherapy in the treatment of panic: A preliminary investigation. *Behavior Therapy, 24,* 241–252.

Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist, 47*, 766–779.

Gregory, R. J., Canning, S. S., Lee, T. W., & Wise, J. D. (2004). Cognitive bibliotherapy for depression: A meta-analysis. *Professional Psychology: Research and Practice, 35,* 275–280.

Gunter, R. W., & Whittal, M. L. (2010). Dissemination of cognitive-behavioral treatments for anxiety disorders: Overcoming barriers and improving patient access. *Clinical Psychology Review*, *30*, 194–202.

Hampson, S. E. (1996). Illness representations and self-management of diabetes. In J. Weinman & K. Petrie (Eds.), *Perceptions of illness and treatment: Current psychological research and applications*. Chur, Switzerland: Harwood Academic.

Heiman, J., LoPiccolo, L., & LoPiccolo, J. (1976). *Becoming orgasmic: A sexual growth program for women*. Englewood Cliffs, NJ: Prentice-Hall.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9,* 1–23.

Jack, S. J., & Ronan, K. R. (2008). Bibliotherapy: Practice and research. *School Psychology International, 29,* 161–182.

Jacobs, M. K., & Goodman, G. (1989). Psychology and self-help groups: Predictions on a partnership. *American Psychologist, 44,* 536–545.

Jacobson, E. (1934). *You must relax: A practical method of reducing the strains of modern living*. New York: McGraw-Hill.

Jamison, C., & Scogin, F. (1995). The outcome of cognitive bibliotherapy with depressed adults. *Journal of Consulting and Clinical Psychology, 63,* 644–650.

Jeffcoat, T., & Hayes, S. C. (2012). A randomized trial of ACT bibliotherapy on the mental health of K–12 teachers and staff. *Behaviour Research and Therapy, 50,* 571–579.

Jerome, L. W., & Zaylor, C. (2000). Cyberspace: Creating a therapeutic environment for telehealth applications. *Professional Psychology: Research and Practice, 31,* 478–483.

Johnson, B. W., & Johnson, W. L. (1998). Self-help books used by religious practitioners. *Journal of Counseling and Development, 76,* 459–466.

Katz, A. H., & Bender, E. I. (Eds.). (1976). *The strength in us: Self-help groups in the modern world*. New York: New Viewpoints.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science, 6,* 21–37.

Kiluk, B. D., Sugarman, D. E., Nich, C., Gibbons, C. J., Martino, S., et al. (2011). A methodological analysis of randomized clinical trials of computer-assisted therapies for psychiatric disorders: Toward improved standards for an emerging field. *American Journal of Psychiatry, 168,* 790–798.

Klein, B., & Richards, J. D. (2001). A brief Internet-based treatment for panic disorder. *Behavioural and Cognitive Psychotherapy, 29,* 113–117.

Koshland, D. (1991). Credibility in science and the press. *Science, 254,* 629.

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science, 16,* 617–625.

Kurtzweil, P. L., Scogin, F., & Rosen, G. M. (1996). A test of the fail-safe N for self-help programs. *Professional Psychology: Research and Practice, 27,* 629–630.

Lanza, M. L. (1996). Bibliotherapy and beyond. *Perspective in Psychiatric Care, 32,* 12–14.

Lazarus, A. (1977). *In the mind's eye*. New York: Rawson.

Leahy, R. L. (2005). *The worry cure: Seven steps to stop worry from stopping you*. New York: Three Rivers Press.

Leventhal, H., & Diefenbach, M. (1991). The active side of illness cognition. In J. A. Skelton & R. T. Croyle (Eds.), *Mental representation in health and illness* (pp. 246–272). New York: Springer-Verlag.

Lewinsohn, P., Munoz, R. F., Zeiss, A., & Youngren, M. A. (1979). *Control your depression*. Englewood Cliffs, NJ: Prentice-Hall.

Lichtenstein, E., Glasgow, R. E., Lando, H. A., Ossip-Klein, D. J., & Boles, S. M. (1996). Telephone counseling for smoking cessation: Rationales and review of evidence. *Health Education Research, 11,* 243–257.

Lidran, D. M., Watkins, P. L., Gould, R. A., & Clum, G. A. (1995). A comparison of bibliotherapy and group therapy in the treatment of panic disorder. *Journal of Consulting and Clinical Psychology, 62,* 865–869.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2,* 53–70.

Lintvedt, O. K., Griffiths, K. M., Sørensen, K., Østvik, A. R., Wang, C. E., Eisemann, M., et al. (2013). Evaluating the effectiveness and efficacy of unguided Internet-based self-help intervention for the prevention of depression: A randomized controlled trial. *Clinical Psychology and Psychotherapy, 20,* 10–27.

Lofflin, J. (1988, March 20). Help from the hidden persuaders. *The New York Times.*

Lowe, J. C., & Mikulas, W. L. (1975). Use of written material in learning self-control of premature ejaculation. *Psychological Reports, 37,* 295–298.

Mahoney, M. J., & Mahoney, K. (1976). *Permanent weight control*. New York: Norton.

Mahoney, M. J., & Thoresen, C. E. (1974). *Self-control: Power to the person.* Monterey, CA: Brooks-Cole.

Marks, I. M., Mataix-Cols, D., Kenwright, K., Cameron, R., Hirsch, S., & Gega. L. (2003). Pragmatic evaluation of computer-aided self help for anxiety and depression. *British Journal of Psychiatry, 183, 57–65.*

Marks, I. M., Shaw, S., & Parkin, R. (1998). Computer-aided treatments of mental health problems. *Clinical Psychology: Science and Practice, 5,* 151–170.

Marrs, R. W. (1995). A meta-analysis of bibliotherapy studies. *American Journal of Community Psychology, 23*, 843–870.

Marshall, W. L., Presse, L., & Andrews, W. R. (1976). A self-administered program for public speaking anxiety. *Behaviour Research and Therapy, 14,* 33–40.

Matson, J. L., & Ollendick. T. H. (1977). Issues in toilet training normal children. *Behavior Therapy, 8,* 549–553.

Merikle, P. M. (1988). Subliminal auditory tapes: An evaluation. *Psychology and Marketing, 46,* 355–372.

Merikle, P. M., & Skanes, H. (1992). Subliminal self-help audiotapes: A search for placebo effects. *Journal of Applied Psychology, 77,* 772–776.

Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist, 24,* 1063–1075.

Mitchell, J. E., Agras, S., Crow, S., Halmi, K., Fairburn, C. G., Bryson, S., et al. (2011). Stepped care and cognitive behavioral therapy for bulimia nervosa: Randomised trial. *British Journal of Psychiatry, 198,* 391–397.

Moore, T. E. (1995). Subliminal self-help auditory tapes: An empirical test of perceptual consequences. *Canadian Journal of Behavioral Science, 27,* 9–20.

Muto, T., Hayes, S. C., & Jeffcoat, T. (2011). The effectiveness of Acceptance and Commitment Therapy bibliotherapy for enhancing the psychological health of Japanese college students living abroad. *Behavior Therapy, 42,* 323–335.

Newman, M. G., Consoli, A., & Talor, C. B. (1997). Computers in assessment

and cognitive behavioral treatment of clinical disorders: Anxiety as a case in point. *Behavior Therapy, 28,* 211–235.

Newsweek (2000, January 10). *Self-help U.S.A.*, pp. 43–47.

Norcross, J. C. (2000). Here comes the self-help revolution in mental health. *Psychotherapy, Research, Practice, Training, 37,* 370–377.

Norcross, J. C., Campbell, L. F., Grohol, J. M., Santrock, J. W., Selagea, F., & Sommer, R. (2013). *Self-help that works: Resources to improve emotional health and strengthen relationships*. New York: Oxford University Press.

Norcross, J. C., Pfund, R. A., & Prochaska, J. O. (2013). Psychotherapy in 2022: A Delphi poll on its future. *Professional Psychology: Research and Practice, 44,* 363–370.

Norcross, J. C., Santrock, J. W., Campbell, L. F., Smith, T. P., Sommer, R., & Zuckerman, E. L. (2000). *Authoritative guide to self-help resources in mental health*. New York: Guilford Press.

Oldenburg, B., Hardcastle, D. M., & Kok, G. (1997). Diffusion of innovations. In K. Glanz, F. M. Lewis, & B. Rimer (Eds.), *Health behavior and education research: Theory, research and practice* (pp. 270–286). San Francisco: Jossey-Bass.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review: A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy, 10,* S21–S39.

Pardeck, J. T. (1990a). Bibliotherapy with abused children. *Families in Society, 71,* 229–235.

Pardeck, J. T. (1990b). Children's literature and child abuse. *Child Welfare, 69,* 83–88.

Pardeck, J. T. (1990c). Using bibliotherapy in clinical practice with children. *Psychological Reports, 67,* 1043–1049.

Pardeck, J. T. (1991a). Bibliotherapy and clinical social work. *Journal of Independent Social Work, 5,* 53–63.

Pardeck, J. T. (1991b). Using books to prevent and treat adolescent chemical dependency. *Adolescence, 26,* 201–208.

Pardeck, J. T. (1992a). Using reading materials with childhood problems. *Psychology: Journal of Human Behavior, 28,* 58–65.

Pardeck, J. T. (1992b). Using books in clinical practice. *Psychotherapy in Private Practice, 9,* 105–119.

Pardeck, J. T. (1993). *Using bibliotherapy in clinical practice: A guide to self-help books*. Westport, CT: Greenwood Press.

Pardeck, J. T. (1994). Using literature to help adolescents cope with problems. *Adolescence, 29,* 421–427.

Pardeck, J. T. (1996). Bibliotherapy: An innovative approach for helping children. *Early Child Development and Care, 110,* 83–88.

Pardeck, J. T. (1997). Recommended self-help books for families experiencing divorce: A specialized form of bibliotherapy. *Psychotherapy in Private Practice, 15,* 45–58.

Pardeck, J. T., & Markward, M. J. (1995). Bibliotherapy: Using books to help children deal with problems. *Early Child Development and Care, 106,* 75–90.

Pardeck, J. T., & Pardeck, J. A. (1993). *Bibliotherapy: A clinical approach for helping children*. Langhorne, PA: Gordon & Breach Science.

Pardeck, J. T., & Pardeck, J. A. (1999). An exploration of the uses of children's

books as an approach for enhancing cultural diversity. *Early Child Development and Care, 147,* 25–31.
Peale, N. V. (1952). *The power of positive thinking.* Englewood Cliffs, NJ: Prentice-Hall.
Phillips, R. E., Johnson, G. D., & Geyer, A. (1972). Self-administered systematic desensitization. *Behaviour Research and Therapy, 10,* 93–96.
Pratkanis, A. (1992). The cargo cult science of subliminal persuasion. *Skeptical Inquirer, 16,* 260–272.
Pratkanis, A. R., Eskanazi, J., & Greenwald, A. G. (1994).What you expect is what you believe (but not necessarily what you get): A test of the effectiveness of subliminal self-help audiotapes. *Basic and Applied Social Psychology, 15,* 251–276.
Quackenbush, R. L. (1992). The prescription of self-help books by psychologists: A bibliography of selected bibliotherapy resources. *Psychotherapy, 28,* 671–677.
Redding, R. E., Herbert, J. D., Forman, E. M., & Gaudiano, B. A. (2008). Popular self-help books for anxiety, depression, and trauma: How scientifically grounded and useful are they? *Professional Psychology: Research and Practice, 39,* 537–545.
Robitscher, J. (1980). *The powers of psychiatry.* Boston: Houghton-Mifflin.
Rosen, G. M. (1976a). The development and use of nonprescription behavior therapies. *American Psychologist, 31,*139–141.
Rosen, G. M. (1976b). *Don't be afraid.* Englewood Cliffs, NJ: Prentice-Hall.
Rosen, G. M. (1977). Nonprescription behavior therapies and other self-help treatments: A reply to Goldiamond. *American Psychologist, 32,* 178–179.
Rosen, G. M. (1987). Self-help treatment books and the commercialization of psychotherapy. *American Psychologist, 42,* 46–51.
Rosen, G. M. (1993). Self-help or hype?: Comments on psychology's failure to advance self-care. *Professional Psychology: Research and Practice, 24,* 340–345.
Rosen, G. M. (1994). Self-Help Task Forces revisited: A reply to Dr. Lowman. *Professional Psychology: Research and Practice, 25,* 100–101.
Rosen, G. M., Glasgow, R. E., & Barrera, M., Jr. (1976). A controlled study to assess the clinical efficacy of totally self-administered systematic desensitization. *Journal of Consulting and Clinical Psychology, 44,* 208–217.
Rosen, G. M., Glasgow, R. E., & Moore, T. E. (2003). Self-help therapy: The science and business of giving psychology away. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 399–424). New York: Guilford Press.
Roy-Byrne, P., Craske, M. G., Sullivan, G., Rose, R. D., Edlund, M. J., et al. (2010). Delivery of evidence-based treatment for multiple anxiety disorders in primary care. *Journal of the American Medical Association, 303*, 1921–1928.
Russell, T. G., Rowe, W., & Smouse, A. (1991). Subliminal self-help tapes and academic achievement: An evaluation. *Journal of Counseling and Development, 69,* 359–362.
Scogin, F., Bynum, J., Stephens, G., & Calhoon, S. (1990). Efficacy of self-administered treatment programs: Meta-analytic review. *Professional Psychology: Research and Practice, 21,* 42–47.
Scogin, F., Jamison, C., & Davis, N. (1990). Two-year follow-up of bibliotherapy for depression in older adults. *Journal of Consulting and Clinical Psychology, 58,* 665–667.

Scogin, F., Jamison, C., & Gochneaur, K. (1989). Comparative efficacy of cognitive and behavioral bibliotherapy for mildly and moderately depressed older adults. *Journal of Consulting and Clinical Psychology, 57,* 403–407.

Singer, J. L., & Switzer, E. (1980). *Mind play: The creative uses of fantasy.* Englewood Cliffs, NJ: Prentice-Hall.

Smiles, S. (1881). *Self-help: With illustrations of character, conduct, and perseverance.* Chicago: Belford, Clarke, & Co.

Smiles, S. (1886). *Happy homes and the hearts that make them.* Chicago: U.S. Publishing House.

Sorge, G .B., Moore, T. E., & Toplak, M. E. (2009). Comparing the content of parenting self-help programs with evidence-based parenting programs: Understanding the scientific grounding of self-help advice. *Scientific Review of Mental Health Practice, 7,* 26–36.

Starker, S. (1989). *Oracle at the supermarket: The American preoccupation with self-help books.* New Brunswick, NJ: Transaction.

Striegel-Moore, R. H., Wilson, G. T., DeBar, L., Perrin, N., Lynch, F., Rosselli, F., et al. (2010). Cognitive behavioral guided self-help for the treatment of recurrent binge eating. *Journal of Consulting and Clinical Psychology*, *78,* 312–321.

Strom, L., Pettersson, R., & Andersson, G. (2000). A controlled trial of self-help treatment of recurrent headache conducted via the Internet. *Journal of Consulting and Clinical Psychology, 68,* 722–727.

Strom, L., Pettersson, R., & Andersson, G. (2004). Internet-based treatment for insomnia: A controlled evaluation. *Journal of Consulting and Clinical Psychology, 72,* 113–120.

Thorsell, J., Finnes, A., Dahl, J., Lundgren, T., Gybrant, M., Gordh, T., et al. (2011). A comparative study of 2 manual-based self-help interventions: Acceptance and commitment therapy and applied relaxation, for persons with chronic pain. *Clinical Journal of Pain, 27,* 716–723.

The Wall Street Journal. (2000, January 7). *Best selling books.* p. W4.

Warner, R. E. (1992). Bibliotherapy: A comparison of the prescription practices of Canadian and American psychologists. *Canadian Psychology, 32,* 529–530.

Watkins, P. L., & Clum, G. A. (2008). *Handbook of self-help therapies.* New York: Taylor & Francis Group.

Winnett, R. A., King, A. C., & Altman, D. G. (1989). *Health psychology and public health: An integrative approach.* New York: Pergamon.

Young, J. H. (1961). *The toadstool millionaires: A social history of patent medicines in America before federal regulation.* Princeton, NJ: Princeton University Press.

Zeiss, R. A. (1977). Self-directed treatment for premature ejaculation: Preliminary case reports. *Journal of Behavior Therapy and Experimental Psychiatry, 8,* 87–91.

Zeiss, R. A. (1978). Self-directed treatment for premature ejaculation. *Journal of Consulting and Clinical Psychology*, *46*, 1234–1241.

Zeiss, R. A., & Zeiss, A. (1978). *Prolong your pleasure.* New York: Pocket Books.

Zimbardo, P. G. (1977). *Shyness.* New York: Jove.

PART III

CONTROVERSIES IN THE TREATMENT OF ADULT DISORDERS

CHAPTER TEN

Science- and Non-Science-Based Treatments for Trauma-Related Stress Disorders

Jeffrey M. Lohr, Richard Gist, Brett Deacon, Grant J. Devilly, and Tracey Varker

In this chapter we critically examine widely promoted interventions for psychological trauma and its sequelae. Because the field of trauma treatment has been fertile ground for treatments with questionable claims of efficacy and effectiveness, careful scrutiny of all trauma treatments is warranted. We begin by discussing the features of psychological trauma and its prevalence. We next describe the two major diagnostic disorders that are reactions to traumatic events: posttraumatic stress disorder (PTSD) and acute stress disorder (ASD). We then address the nature of treatment efficacy research on trauma and its relation to treatment effectiveness. In doing so, we outline contemporary cognitive-behavioral theories of anxiety disorders and describe empirically supported psychosocial treatments for those conditions. We next discuss promising psychosocial procedures for pre-event and postevent prevention of the disorders. Finally, we discuss the professional and social implications of the implementation of science-based interventions for trauma.

TRAUMA AND ITS CONSEQUENCES

The most frequent traumas are the violent death of a loved one, robbery, physical or sexual assault, motor vehicle accident, or natural disaster

(Breslau et al., 1998; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). Epidemiological evidence indicates that between 50 and 70% of adults have experienced at least one such event in their lifetime (Kessler et al., 1995). The bulk of treatment efficacy research to be reviewed has been predicated on the formal diagnosis of PTSD first identified in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III, DSM-III-R; American Psychiatric Association, 1980, 1987) and a subsequent revision that included ASD (DSM-IV; American Psychiatric Association, 1994). In the context of PTSD and ASD diagnoses, trauma has been defined as (1) experiencing, observing, or confronting an event that has involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others; and (2) where the person has responded with intense fear, helplessness, or horror (American Psychiatric Association, 1994).

One potential consequence of traumatic experience is PTSD, which in DSM-IV (American Psychiatric Association, 1994) consists of the following core symptoms:

1. Mentally reexperiencing the traumatic event. Such symptoms include recurrent and intrusive distressing recollections of the event, recurrent distressing dreams of the event, acting or feeling as if the event were happening again, or psychological or physiological distress when exposed to stimuli that remind the person of the traumatic event.
2. Avoidance of stimuli associated with the trauma, or numbing of general responsiveness. Symptoms in this category include efforts to avoid thoughts, feelings, or conversations associated with the trauma; efforts to avoid activities, places, or people that arouse recollections of the trauma; inability to recall an important aspect of the trauma; diminished interest in usual activities; feelings of detachment from others; restricted range of affect; and sense of a foreshortened future.
3. Increased arousal. Symptoms include sleep disturbance, irritability or anger outbursts, difficulty in concentration, hypervigilance, and exaggerated startle response. In addition, the formal diagnosis requires that the symptoms must last at least one month and cause significant distress or functional impairment.

The fifth edition of the DSM (DSM-5; American Psychiatric Association, 2013) includes a number of modifications in the diagnosis of PTSD. In addition to the domains already described, a fourth domain involving negative alterations in cognition and mood is required. Several other specific details have also been altered. First, the requirement that a person experience intense emotion at the time of the trauma has been removed. Second, the criteria for trauma are more narrow and specific, such that mere

media exposure is no longer considered traumatic. Third, either avoidance symptoms or numbing symptoms are required as they appear to be distinct phenomena (Asmundson, Stapleton, & Taylor, 2004). Fourth and finally, symptoms involving concentration, sleep disturbance, and diminished pleasurable interest must begin after the trauma.

In DSM-5 (American Psychiatric Association, 2013), ASD symptoms include dissociative experiences during or following the event, reexperiencing symptoms, avoidance symptoms, and arousal symptoms. The symptoms are experienced during the event or immediately thereafter, last for at least 2 days or resolve within 4 weeks of the event; if they do not, the diagnosis is changed to another disorder (e.g., PTSD, major depressive disorder). In DSM-5, the onset is 3 to 31 days posttrauma. Dissociation is no longer a required symptom, but two dissociative symptoms are included. The number of symptoms has increased from 8 to 14, and 8 of those 14 are required for diagnosis. The symptoms are now more similar in content to PTSD (Kring, Johnson, Davison, & Neale, 2012).

Contemporary studies of the general population indicate a lifetime prevalence for PTSD ranging from 6.8% (Kessler et al., 2005) to 7.8% (Kessler et al., 1995). The conditional risk of developing PTSD following a trauma varies greatly, depending on the nature of the event. For males, having engaged in combat and witnessing violence are most likely to lead to PTSD. For females, rape and sexual molestation are most likely to lead to PTSD (Kessler et al., 1995). The prevalence of ASD depends on the degree of exposure, the persistence of the exposure, and the severity of the event (American Psychiatric Association, 1994, 2013).

SCIENCE-BASED TREATMENTS COMPARED WITH "EVIDENCE-BASED" TREATMENTS

Comparisons

The phrase "science-based treatments" comports much more directly with experimental methods than does the phrase "evidence-based treatment" because the contemporary definition of evidence in clinical psychology is typically broader and more diffuse, encompassing not merely research support but also clinical expertise and client preferences and values. The former phrase was coined in the context of the attempt by the American Psychological Association's Division 12 (Society for Clinical Psychology) to develop criteria for identifying empirically supported treatments (Task Force on Promotion and Dissemination of Psychological Procedures, 1995). The result was a "box-score" approach to scientific legitimacy (Herbert, 2000, 2003) that led to a decade of professional debates regarding list qualifications, list status, training implications, and reimbursement concerns. Tasked with addressing such issues, the American Psychological Association Presidential Task Force on Evidence-Based Practice (American

Psychological Association, 2006) issued a document that codified the nature of "evidence" for evaluating treatment efficacy, effectiveness, and justification for clinical application. In response, Stuart and Lilienfeld (2007) argued that the conceptualization of "evidence-based practice" arrived at by the Presidential Task Force was largely devoid of an empirical foundation and broadened the domain of "evidence" to include virtually any and all forms of clinical opinion. Thus, we are left with the need to better specify the nature of the empirical evidence for treatment efficacy and the theories from which those treatments derive.

The theory of a treatment should also be considered when making judgments about the validity of a particular treatment (David & Montgomery, 2011; Lilienfeld, 2011). Specifically, a more direct consideration of how the relationship of the mechanisms of treatment relate to the mechanisms of the disorder being treated is needed. In the best of all scientific worlds, there should be a correspondence between the mechanisms of treatment (the active treatment ingredients) and the maintenance mechanisms of the disorder (the treatment targets). Ideally, methods to ascertain a treatment's efficacy and effectiveness should compare treatments with credible nonspecific control conditions to ascertain the additive effect of the specific content over, if any, and above therapeutic alliance and other common factors. However, strong tests of the incidental and specific content of any prescriptive, structured (or manualized) treatment should be the goal of the evaluative framework for clinical efficacy. This goal can be accomplished using component–control experimental designs (often called "dismantling designs") to test whether the specific content of the treatment functions as the "active ingredient" responsible for apparent efficacy (Borkovec & Castonguay, 1998; Kazdin, 2002). Treatments can be considered "specific" if such experiments reveal that the necessary and specific components provide a meaningful increment in efficacy beyond incidental aspects of treatment. Such methodological concerns apply most directly to novel or recently introduced treatments that are promoted as distinct, specific, innovative, and efficacious.

Appeal to "evidence" by proponents of these treatments will be helpful only if the methodological rigor of outcome experiments undergirds that evidence. The experimental rigor of outcome studies will be maximized when the theoretical mechanisms of the treatments are submitted to strong evidentiary tests, as we have outlined. The strength of the evidence will be maximized further if the theoretical mechanisms of the disorder inform the theoretical mechanisms of the treatments. Only when we have experimental evidence that the specific content of a treatment results in change of the specific features of a disorder can we claim, in our view, to have developed empirically supported psychosocial interventions for psychological disorders. In this respect, our position differs from that of the Division 12 committee that outlined criteria for empirically supported interventions that focus exclusively on the question of how well the treatment works. In doing so, we will also be in a better position to identify treatments that

are actively harmful or are inert. Such analyses also may have the salutary effect of focusing our evaluative efforts on evidence-based principles of psychosocial intervention rather than on named interventions (see Rosen & Davison, 2003).

Clinical Science: Demonstration of Efficacy and Effectiveness

Elucidation of active treatment components is achieved through the experimental analysis of treatment efficacy. In this analysis, (1) a treatment's efficacy is tested against a logical progression of comparison groups, and (2) the treatment package is broken down into its constituent components that are independently tested to determine which, if any, are therapeutically active (Borkovec & Castonguay, 1998; Lohr, DeMaio, & McGlynn, 2003; Lohr, Olatunji, Parker, & DeMaio, 2005). The strength of the evidence for treatment efficacy increases monotonically when evaluated against the following comparison conditions: wait-list control, attention control, nonspecific factor control, alternative credible treatment, and working alliance (Wampold, 2001). This progressive dismantling strategy demonstrates that the procedure contains an active ingredient that adds incremental efficacy to the treatment beyond that attributable to nonspecific factors (Lohr et al., 2003, 2005). Treatment effectiveness first presumes treatment efficacy (see Gaudiano, Dalrymple, Weinstock, & Lohr, Chapter 6, this volume). Once efficacy has been demonstrated, evaluating effectiveness includes assessing the generality of the treatment's clinical application, feasibility of implementation, cost-effectiveness, and clinical impact.

The gap between mental health practice and the science that provides the evidential warrant for interventions has grown and may be accelerating in some quarters (Lilienfeld, Lynn, & Lohr, Chapter 1, this volume). The manifestations of this widening disjunction are most obvious in the public domain, where clinical services are often promoted with little scientific justification (Gambrill, 2012). The introduction of certain treatments claimed by proponents to be both novel and extraordinary has often been based on little more than personal testimony and vivid case studies (Herbert et al., 2000; Kalal, 1999; Lohr, 2001).

TRAUMA TREATMENT AS OBJECT LESSON

Pseudoscientific Treatment

The field of "traumatology" has rapidly expanded since PTSD was first introduced in DSM-III as an anxiety disorder often resulting from armed combat (American Psychiatric Association, 1980). The rapid expansion of the traumatology field has occurred because of the pain and suffering that often follow trauma, the apparent refractory nature of PTSD, and the often compensable consequences of the diagnosis.

In the domain of mental health practice, pseudoscience is typically

marked by confirmation bias and inattention to disconfirming data (Herbert et al., 2000). In the place of empirical evidence, pseudoscientific practice often reverts to personal testimony and anecdotal evidence (Gaudiano & Herbert, 2000). Another characteristic of clinical pseudoscience is the failure to acknowledge boundary conditions of theory and limitations of application. Treatments that are based on science and theory typically are directed at specific problems. In contrast, treatments that are based on pseudoscience are often promoted as panaceas with no limits and are said to be applicable to the widest range of clinical problems (Hines, 2003). Pseudoscience is also commonly characterized by an overpromotion of interventions to the general public relative to the available scientific evidence. Trauma treatments based on pseudoscience are often popularized in print and broadcast media long before they are subjected to controlled tests (Gaudiano & Herbert, 2000; Herbert et al., 2000).

The proliferation of pseudoscience in the mental health professions may be attributable to the resistance of many mental health practitioners to explicitly establishing scientific criteria for treatment efficacy (e.g., Fox, 2000). This reluctance appears to stem from four fundamental concerns: (1) the health care system would be faced with the responsibility of ensuring that treatments meet those criteria; (2) scientifically explicit criteria would reduce the flexibility by which treatments are provided and would likely delay the entry of new treatments into the profession; (3) scientifically explicit criteria would force many clinicians to forgo favored treatments for which research evidence is lacking; and (4) many clinicians would incur a professional obligation to be (re)trained in empirically supported treatments (ESTs); (Beutler, 2000; Nathan, 2000). The implementation of scientific criteria for evaluating treatments will surely not be without its difficulties (e.g., Herbert, 2000, 2003). However, the continued tolerance of pseudoscience in trauma-related mental health practice places the public at risk and may violate the public's trust in the profession (Beyerstein, 2001; Devilly & Lohr, 2008; Newbold, Lohr, & Gist, 2008).

Promotion of pseudoscience may also be accompanied by promotions of "junk science." Junk science comes in three major forms. One occurs in the context of legal proceedings, entering as unsubstantiated "expert" testimony by psychologists and other mental health professionals in the context of litigation (Huber, 1991; Park, 2000; see McCann, Lynn, Lilienfeld, Shindler, & Hammond, Chapter 4, this volume). Such testimony addresses psychological phenomena or services for which scientific evidence does not meet legal standards (e.g., federal rules of evidence), as in the *Frye* test or the *Daubert* standard (see Chapter 4, this volume). The second kind of junk science is sometimes found in the media, where mental health issues are commercialized through entertainment, advertising, and psychological advice-giving (Wilson, 2003). The third form of junk science occurs in the promotion of services to consumers and other parties (e.g., governmental agencies) with a stake in the efficacy of such services. In each of these contexts, junk science is manifested by opinion posing as empirical evidence,

or through evidence of questionable warrant, based on inadequate scientific methodology (Devilly & Lohr, 2008; Newbold et al., 2008). The following sections feature examples of such questionable practices in the development of the field of traumatology since the introduction of trauma-related diagnoses (American Psychiatric Association, 1980, 1987, 1994, 2013).

Eye Movement Desensitization and Reprocessing

Treatment Description and Rationale

A treatment developed by Francine Shapiro (1989) known as eye movement desensitization and reprocessing (EMDR) has become remarkably popular in recent decades. The idea of EMDR was born during a serendipitous walk in the woods in 1987 (Shapiro, 1995a). While thinking of an anxiety-provoking situation, Shapiro noticed that her eyes were involuntarily moving back and forth, and that her distressing thoughts subsequently disappeared. After a series of successful trials in her private practice and an uncontrolled study demonstrating the effectiveness of a single session of eye movements with 22 clients distressed by traumatic memories (Shapiro, 1989), Shapiro began training clinicians in her approach. Since 1992, more than 70,000 clinicians have received training in EMDR by attending proprietary certification workshops conducted by Shapiro's EMDR Institute (Shapiro, 2005). The EMDR International Association (EMDRIA) holds numerous training and research conferences and promotes EMDR within the mental health field. EMDR has been widely advertised to the public, mental health consumers, and therapists using a variety of debatable, yet apparently effective, marketing tactics (Herbert et al., 2000). As a result, EMDR has enjoyed rapid and widespread enthusiasm among practitioners. To illustrate, the number of entrants in the EMDRIA's therapist directory dwarfs the membership of therapist directories associated with other anxiety- and PTSD-related specialist organizations, including the Anxiety Disorders Association of America, the Association of Behavioral and Cognitive Therapies, and the International Society for Traumatic Stress Studies. As described further in this section, EMDR is distinctive among evidence-based PTSD treatments by virtue of its novel theoretical model and treatment procedures as well as its embodiment of numerous features of pseudoscience as discussed in Lilienfeld, Lynn, and Lohr, Chapter 1, this volume.

EMDR uses a structured, prescriptive intervention procedure that incorporates such general clinical components as history taking and verbal report of the nature and emotional consequences of the traumatic experience. In addition, the EMDR procedure requires the client to construct and maintain both an imaginal representation or a memory (or other image) and the physical sensations associated with the traumatic event. While maintaining the image, the therapist introduces one or more "bilateral stimulation" procedures, such as moving a finger across the client's visual field to produce side-to-side eye movements. Such procedures are intended

to elicit simultaneous "dual attention" to internal and external stimuli. The client is asked to express the negative cognitions that accompany the affective distress and to generate a more positive appraisal about the trauma and his or her experience with it. This component is referred to as "reprocessing" and is added to the desensitization that accompanies the imaginal exposure (Shapiro, 2001).

EMDR is based on a set of theoretical conjectures that rely heavily on physiological concepts related closely to neurological processes. The nature of trauma pathology and its effective treatment is predicated on a model called accelerated information processing (AIP), which is ostensibly akin to a psychological immune system (Shapiro, 1995a). Healing is posited to occur after eye movements and other features of the clinical protocol presumably "unlock" the pathological condition. The AIP model defines pathology as "dysfunctionally stored information that can be properly assimilated through a dynamically activated processing system" (Shapiro, 1995a, p. 52). Based on this formulation, the practice of EMDR involves:

> accessing the dysfunctionally stored information, stimulating the innate processing system through the standardized protocols and procedures (including the bilateral stimulation), and facilitating dynamic linkages to adaptive memory networks, thereby allowing the characteristics of the memory to change as it transmutes to an adaptive resolution. (Solomon & Shapiro, 2008, p. 316)

Despite the superficial appeal of this neurophysiological speculation, the AIP model has little scientific basis and relies heavily on obscurantist language to create the appearance of scientific legitimacy. Keane (1998) highlighted the absence of connectivity between the theoretical foundations of EMDR and scientific knowledge on the nature and treatment of PTSD (see Stanovich, 2012). Specifically, the AIP model and bilateral stimulation techniques are largely divorced from existing models of psychopathology and psychotherapy, and are largely inconsistent with the body of scientific knowledge gleaned from experimental psychopathology regarding the nature, acquisition, and modification of fear and anxiety.

Comparison with Other Evidence-Based Psychological Treatments

Early reviews of the EMDR outcome literature (Cahill, Carrigan, & Frueh, 1999; DeBell & Jones, 1997; Herbert et al., 2000; Lohr, Lilienfeld, Tolin, & Herbert, 1999; Lohr, Tolin, & Lilienfeld, 1998) questioned the efficacy of this therapy in light of methodological limitations in studies that purported to show clinical effects. The evidence base for EMDR has increased in quantity and quality since the first edition of this book (Lilienfeld, Lynn, & Lohr, 2003), and this therapy is now regarded as efficacious in the treatment of PTSD in several clinical practice guidelines (American Psychiatric Association, 2004; Australian Centre for Posttraumatic Mental Health,

2007; Department of Veterans Affairs/Department of Defense, 2010; National Institute for Clinical Excellence, 2005).

Two well-conducted clinical trials have directly compared the efficacy of EMDR and prolonged exposure (PE), which is a widely researched, empirically supported treatment for PTSD (Foa, Gillihan, & Bryant, 2013). Taylor et al. (2003) randomly assigned 60 participants with PSTD to receive eight 90-minute sessions of EMDR, exposure therapy (both imaginal and *in vivo*), or relaxation training. Self-report and blind assessor ratings of PTSD symptoms were gathered at pretreatment, 1-month posttreatment, and 3-month follow-up. Treatment conditions did not differ significantly with respect to attrition and credibility, and fidelity to the protocol was high for each intervention. Findings indicated that all three treatments were effective in reducing PTSD symptoms, guilt, anger, and depression at posttreatment and follow-up. Compared with EMDR, exposure therapy yielded a significantly higher percentage of participants with clinically significant change and greater reductions in avoidance and reexperiencing symptoms. EMDR and relaxation did not differ significantly in the magnitude or speed of improvement in PTSD symptoms.

A second methodologically rigorous study by Rothbaum, Astin, and Marsteller (2005) compared the efficacy of EMDR, PE, and a wait-list control condition in a sample of 74 adult female rape victims with PTSD. Participants assigned to active treatment received nine twice-weekly sessions of EMDR or PE. Twenty participants completed each treatment; attrition was low and did not differ significantly between conditions. Blind assessments were conducted at pretreatment, posttreatment, and 6-month follow-up. Both active treatments produced significantly greater improvement in self-reported and clinician-rated PTSD symptoms at posttreatment than did the wait-list condition and did not differ significantly from each other. At posttreatment, the percentages of patients who continued to meet the criteria for PTSD were 5%, 25%, and 90% for PE, EMDR, and wait list, respectively. Significantly more patients who received PE demonstrated good end-state functioning (78%) than those who received EMDR (35.3%) at 6-month follow-up. However, interpretation of this finding is complicated by the fact that the EMDR group evidenced higher scores on several psychopathology measures at baseline.

Several additional clinical trials have compared the efficacy of EMDR and trauma-focused CBT. Power, McGoldrick, Brown, Buchanan, Sharp, et al. (2002) randomly assigned 105 Scottish clients with PTSD to receive a maximum of 10 sessions of EMDR or exposure plus cognitive restructuring (E+CR) in a primary care setting. A wait-list control condition was also employed. Blind raters assessed efficacy at the end of the 10-week treatment period, and the Clinician Administered PTSD Scale (CAPS) was administered at 15-month follow-up by therapists who were not blind to treatment condition. Both active treatments improved significantly more than did the wait-list condition on PTSD measures and did not differ significantly from each other. Dropout rates were comparable across conditions, and

clinically significant change was achieved by 60% of EMDR participants and 50% of E+CR participants.

Lee, Gavriel, Drummond, Richards, and Greenwald (2002) randomly assigned 24 patients with PTSD to receive seven sessions of EMDR or stress inoculation training with PE (SIT+PE), the latter of which consisted of prolonged imaginal exposure, combined with arousal-reduction techniques and cognitive coping skills. There were no significant differences at posttreatment on self- and observer-rated indices of PTSD and related symptoms, and 83% of EMDR participants and 75% of SIT+PE participants no longer met diagnostic criteria for PTSD. At 3-month follow-up, EMDR was associated with significantly greater improvement on all measures, and clinically significant improvement was evident among 92% of clients who received EMDR and 50% of clients who received SIT+PE.

The aforementioned clinical trials, as well as additional studies using less methodologically rigorous designs (e.g., Devilly & Spence, 1999; Ironson, Freund, Strauss, & Williams, 2002; Tarrier et al., 1999a; Tarrier, Sommerfield, Pilgrim, & Humphreys, 1999b), suggest that the efficacy of EMDR is generally comparable to that of trauma-focused CBT approaches such as PE in the treatment of PTSD (see Spates, Koch, Cusack, Pagoto, & Waller, 2009, for a review). This conclusion has been reached by the authors of multiple meta-analytic reviews published in recent years (Bisson & Andrew, 2007; Bisson et al., 2007; Bradley, Green, Russ, Dutra, & Westen, 2005; Powers, Halpern, Ferenschak, Gillihan, & Foa, 2010; Seidler & Wagner, 2006). Although head-to-head comparisons with alternative therapies have demonstrated that EMDR is efficacious in the treatment of PTSD, evidence supporting the overall therapeutic benefit of this multifaceted therapy is insufficient to demonstrate the validity of the treatment's theoretical rationale. Given the centrality of dual attention procedures in the EMDR protocol (Shapiro, 2001), as well as the novelty and scientifically questionable status of the AIP model (Devilly, 2002), it is important to consider findings from clinical studies examining the incremental contribution of bilateral stimulation techniques to the efficacy of EMDR.

EMDR Component Controls

Although EMDR involves a variety of therapeutic procedures, bilateral stimulation techniques are the distinctive and characteristic feature of this treatment and are emphasized in the theoretical conjectures underlying EMDR's purported benefits (Solomon & Shapiro, 2008). Component–control experiments are useful in determining the extent to which the putative active therapeutic ingredients in EMDR are specifically efficacious and provide a meaningful benefit beyond the nonspecific aspects of treatment. Nonspecific factors in an experimental treatment procedure include the incidental effects of treatment, such as measurement reactivity, regression to the mean, credibility, expectation for improvement, experimental demand,

therapist–experimenter enthusiasm, and therapist–experimenter allegiance. The most important experimental controls for rigorous tests of treatment components rely on additive and subtractive experimental designs (Cahill et al., 1999; Mahoney, 1978; Nezu, 1986; Nezu & Perri, 1989). These "dismantling" designs, discussed earlier, are necessary to identify the components that are specific to EMDR, and to provide a stringent test of the incremental contribution of eye movements and related techniques to the treatment of PTSD.

Renfrey and Spates (1994) recruited 23 trauma victims, 21 of whom met diagnostic criteria for PTSD. Participants were randomly assigned to one of three conditions: standard EMDR; an EMDR analogue in which eye movements were induced by an optical device alternating the position of a light in the right and left peripheral visual field; and an EMDR analogue in which a light blinked in the center of the visual field. Dependent variables included subjective ratings of discomfort, changes in heart rate, and standardized measures of PTSD symptoms. After treatment, 5 of the 23 participants met criteria for PTSD and were roughly evenly distributed across treatment groups. Analyses of heart rate and subjective ratings revealed significant main effects for repeated assessment but no interaction between assessment and treatment condition. Analyses of the standardized measures were not reported. Thus, it appears that the general EMDR procedure, rather than saccadic eye movements per se, was responsible for reductions in self-report indices and heart rate. However, the control conditions did not directly control for measurement reactivity or nonspecific treatment factors.

Boudewyns and Hyer (1996) compared EMDR with an eye constant (EC) imagery analogue (EC) and a no-imagery control (C) procedure in the treatment of combat-related PTSD. All participants received eight sessions of the standard inpatient or outpatient treatment program at a Veterans Administration hospital. Participants in the EMDR and EC groups received between five and eight sessions of EMDR. The EC participants did not engage in eye movements during individual treatment but kept their eyes closed and engaged in imaginal exposure for the same period of time. Participants in the C condition received only the standard group treatment. The analyses revealed that the EMDR and EC conditions showed greater improvement than did the C condition on subjective distress ratings, Profile of Mood States Anxiety scores, and heart rate. The EMDR and EC conditions did not differ significantly from each other. In addition, the three groups demonstrated statistically equivalent change on the CAPS and Impact of Event Scale. Thus, neither eye movements nor any lateral stimulation was necessary for change, suggesting that imagery exposure may be sufficient for change on some indices of PTSD.

Pitman et al. (1996) used a crossover design in which combat-related PTSD clients were randomly assigned to one of two treatment sequences using EMDR or a no-movement imagery analogue (fixed-eye) treatment.

The analogue control procedure consisted of all EMDR components, including movement of the therapist's hand. The participant maintained eye fixation and tapped one finger to correspond to therapist hand movement. Each treatment was applied for a maximum of six sessions once per week. Analyses of variance between treatment conditions revealed no significant differences on psychophysiological measures. On PTSD-related outcome variables, there was limited change (only three of eight measures) within each of the procedures. Use of the control procedure suggests that eye movements confer no clear advantage over other forms of stimulation. Macklin et al. (2000) later reported that all of the participants exposed to EMDR had returned to pretreatment levels of PTSD symptoms 5 years after completion of treatment.

Devilly, Spence, and Rapee (1998) compared EMDR with a no-movement imagery analogue condition that was presented to war veterans with PTSD as "reactive eye dilation desensitization and reprocessing." This procedure involved the full EMDR protocol except that a flashing light was substituted for lateral eye movements. Both treatments were compared with a no treatment control condition that included the same assessment battery as the treatment conditions. Treatment outcome measures included standardized anxiety, depression, and PTSD scales, as well as heart rate and blood pressure. The results showed that both treatment groups improved by posttreatment, but that there was no significant difference between conditions. Participants in the two treatment conditions did not differ significantly on standardized measures from the control condition, but improved more than the control condition when the reliable change index on the Mississippi–PTSD scale was examined. Nevertheless, there was no statistical or clinical difference in symptoms from pretreatment to 6-month follow-up. The authors concluded that eye movements are not the agent of change and that other nonspecific factors are responsible for the high levels of efficacy reported in previous EMDR research conducted without adequate procedural controls.

A number of additional studies have used dismantling designs to examine the incremental validity of eye movements in analogue samples or non-PTSD clinical populations (e.g., Cusack & Spates, 1999; Wilson, Silver, Covi, & Foster, 1996). Devilly (2002) reported that of the 13 published dismantling studies, 11 found no significant benefit to eye movements in EMDR, and the two seemingly positive studies suffered from serious methodological flaws. Although Shapiro (1989) initially claimed that eye movements were necessary for EMDR's efficacy, she subsequently broadened the range of acceptable bilateral stimulation stimuli to include tapping, tactile stimulation, and auditory tones (Shapiro, 1996). This modification further blurred the question of the purported mechanism of action of bilateral stimulation techniques which initially relied on the similarity between EMDR eye movements and those of REM sleep (Shapiro, 1989). To date, EMDR proponents have not specified the neurophysiological mechanisms

that make auditory and tactile bilateral stimulation techniques functionally equivalent to saccadic eye movements in activating the putative curative phenomenon of accelerated information processing. Devilly (2002) observed that by abandoning her claim that eye movements were necessary, Shapiro (1996), in effect, discounted the research investigating the role of eye movements in EMDR by rendering negative component control studies unable to falsify the AIP model.

Numerous analogue studies have examined the effects of bilateral stimulation procedures on psychological and physiological variables. Van den Hout and Englehard (2012) reviewed this experimental research in an article titled "How Does EMDR Work?" The typical study they reviewed investigated the effects of briefly recalling an unpleasant memory, versus recall plus eye movements, in healthy volunteers. The authors argued that eye movements (but not bilateral auditory stimuli) reduce the vividness and emotionality of memories by taxing working memory during recall. This notion contradicts the popular "interhemispheric communication" theory of dual stimulation techniques (Solomon & Shapiro, 2008) by regarding eye movements as a distraction that interferes with the ability to retrieve a memory. Although van den Hout and Englehard claimed that the efficacy of EMDR is attributable to the effects of eye movements on working memory, a more cautious interpretation is warranted. Findings from analogue studies employing brief exposure to unpleasant memories in healthy volunteers are of questionable generalizability to the longer-term treatment of individuals diagnosed with PTSD. Reductions in emotionality and vividness of unpleasant memories elicited by eye movements might temporarily reduce distress in a healthy volunteer, but inhibit longer-term emotional processing and fear in a patient with PTSD. Such a distraction hypothesis has been previously forwarded (Devilly, 2001a) and explicated (Devilly, 2001b) in the case of EMDR. This perspective argues that poorer long-term reductions have been noted, and would be expected, in those with clinical status through an impaired extinction of the fear response. However, within a nonclinical population there is no substantive reason to expect distraction to exert long-term clinical effects.

Of greater importance, however, is that the effects of eye movements are of little clinical relevance in the absence of compelling evidence to support the specific efficacy of bilateral stimulation techniques. These conclusions were reached by Davidson and Parker (2001) in a meta-analysis of 13 component-controlled efficacy studies. They found an effect size not different from zero when comparing standard EMDR without bilateral stimulation. A subsequent analysis of 15 component-controlled studies by Lee and Cuipers (2013) reported a statistically significant effect size of 0.27. They incorporated all visual modalities in their component controls, including therapist-driven bilateral eye movements, concatenate eye movements, negative imagery, and autobiographical memories containing distress to their inclusion criteria. The reported effect size implies that all studies revealed

the same effect, when in fact five studies showed an effect size not different from, or less than, zero (Carrigan & Levis, 1999; Foley & Spates, 1995; Lytle, Hazlett-Stevens, & Borkovec, 2002; Renfrey & Spates, 1994; Sanderson & Carpenter, 1992).

Moreover, several statistical procedures in this meta-analysis are questionable. The analysis used the number of participants, rather than the number of studies analyzed, as the sample size. Such a procedure excessively constricts confidence intervals (Sanchez-Meca & Marin-Martinez, 2008). Although the authors claim that this tactic results in producing only minor differences in outcome, the substantive analytical problem is that the error margin for statistical significance in this meta-analysis is exceptionally narrow. Further, combining effect sizes from multiple domains within one study and the comprehensive meta-analysis software used by these authors decreases the standard error and inflates the effect sizes.

Another concern was the selection of dependent variables. Meta-analyses usually combine dependent variables that possess concurrent validity from one domain. In addition, they do not combine process variables such as in-session subjective units of distress (SUDs) and validity of cognitions (VOC) variables with outcome variables that assess the diagnostic features of psychopathology. The combination of these dependent variables within each study is likely to decrease the standard error surrounding the derived effect size, as the sample size is counted for each dependent variable independently. The combination of these variables also results in the assessment of nonspecific treatment outcomes. Accordingly, we found that when the authors removed SUDs and other process variables from the meta-analysis they found the lower 95% confidence interval of the analysis only .07 above an effect size of zero, which calls into question the robustness of their claims.

Finally, the studies included appeared to lack consistency in the independent variable (eye movements) being tested. Both meta-analyses (Davidson & Parker, 2001; Lee & Cuipers, 2013) included 10 component-controlled studies (Boudewyns, Stwertka, Hyer, Albrecht, & Sperr, 1993; Carrigan & Levis, 1999; Devilly et al., 1998; Dunn, Schwartz, Hatfield, & Wiegele, 1996; Feske & Goldstein, 1997; Foley & Spates, 1995; Gosselin & Mathews, 1995; Renfrey & Spates, 1994; Sanderson & Carpenter, 1992; Wilson et al., 1996). Lee and Cuipers added three studies published after Davidson and Parker's (2001) review (Lee & Drummond, 2008; Lytle et al., 2002; Schubert, Lee, & Drummond, 2011). However, they also included Shapiro (1989), even though that study did not include a component–control design. Shapiro (1989) noted that the control condition "provided a modified flooding procedure" (p. 206). Lee and Cuipers exclude three of the studies included by Davidson and Parker (Merckelbach, Hogervorst, Kampman, & deJong, 1994; Pitman, Orr, Altman, & Longpre, 1996; Tallis & Smith, 1994). In addition, they did not include two studies that employed imaginal exposure and experimentally manipulated

eye movements (Devilly & Spence, 1999; Taylor et al., 2003). Moreover, they included Lee and Drummond (2008)'s results twice in the meta-analysis. A more convincing meta-analysis would have included the 10 studies analyzed by both Davidson and Parker and by Lee and Cuipers, as well as other bona fide component–control design studies published before or after Davidson and Parker (Devilly & Spence, 1999; Lee & Drummond, 2008; Lytle et al., 2002; Merckelbach et al., 1994; Pittman et al., 1986; Schubert et al., 2011; Tallis & Smith, 1994; Taylor et al., 2003), and excluded the study by Shapiro (1989).

Because of these meta-analytic discrepancies, the finding of even a small effect size for eye movements is questionable. Beyond these methodological concerns, clinical judgments of treatment content made on the basis of meta-analytic findings can be widely discrepant from those findings of well-controlled and adequately powered randomized controlled trials (RCTs) (Klein, 2000; LeLorier, Gregoire, Benhaddad, LaPierre, & Derderian, 1997). The discrepancy among the meta-analyses should be considered in light of an early commentary by Kazrin, Durac, and Agteros (1979) on the misuses of meta-analysis. They described a statistical process that they called meta-meta-analysis, stating "the strength of meta-meta-analysis, like the unfortunate person stranded on a desert island, is that it will look at anything" (p. 397). Kazrin et al. whimsically argued how this process, directed at psychotherapy efficacy, could use 850,000 therapists and 2,000,000 research participants and yet fail to reach interpretable conclusions. We believe that such tongue-in-cheek warnings regarding the misuse of meta-analysis should be taken seriously. Indeed, Barlow (2010b) has observed that, when conducting meta-analysis, "it is easy to sit back and 'pick-off' any new study and conclude that it is not 'perfect' " (p. 16).

There exists a clear scientific consensus that bilateral stimulation procedures do not contribute to the efficacy of EMDR (Spates et al., 2009). The American Psychiatric Association's (2004) PTSD guideline states that dismantling studies "show no incremental effect from the use of eye movement or other proxies during the treatment sessions" (p. 59) and "these studies call into question EMDR's theoretical rationale" (p. 59). Clinical practice guidelines from the Department of Veterans Affairs/Department of Defense (2010) assert that "the data do not suggest that eye movements or other forms of kinesthetic stimulation are necessary" (p. 130).

The responses of some EMDR proponents to criticisms of the specific efficacy of bilateral stimulation techniques illustrate several features of pseudoscience described in Lilienfeld, Lynn, and Lohr, Chapter 1, this volume. Principal among these features is the overuse of ad hoc explanations designed to immunize claims of falsification. As detailed by Herbert et al. (2000), examples include (1) claiming that negative studies are invalid because the researchers did not receive EMDR-sanctioned training (Shapiro, 1995a), despite the absence of evidence supporting the specific benefits of such training (Rosen, 1999); (2) claiming that Level II EMDR training

was necessary following negative studies conducted by investigators who had received only Level I training (Shapiro, 1995b); (3) arguing that component–control procedures are actually heretofore unknown variants of EMDR (Shapiro, 1995a); and (4) questioning the competence, motives, and scientific integrity of EMDR critics (Russell, 2008; Shapiro, 1995b).

Some EMDR proponents have further characterized the findings of dismantling studies as "inconclusive," as a result of methodological problems (e.g., Shapiro, 2002) and have argued that firm conclusions that bilateral stimulation techniques lack specific efficacy in EMDR are unjustified (Perkins & Rouanzoin, 2002) and reflect "confirmatory bias" on the part of EMDR critics (Russell, 2008). This perspective reveals a reversed burden of proof in which a claim can be assumed to be true until it has been proven false beyond a reasonable doubt. Although the burden of proof in science rests on the claimant (McFall, 1991), Shapiro and colleagues have not produced convincing evidence in support of the claim that eye movements and related techniques are specifically efficacious for PTSD. Rather, studies that fail to support the purported therapeutic benefits of bilateral stimulation in EMDR are dismissed for various reasons, and finger movement, auditory tones, and kinesthetic tapping remain integral to the theory and practice of EMDR. As such, EMDR's "reliance on a nonfalsifiable neurobiological model of EMDR that is continually modified in the context of disparate research findings" (Russell, 2008, pp. 1740–1741) continues to represent a barrier to the scientific legitimacy of this treatment approach (Devilly, 2002).

EMDR as a Novel and Distinct Treatment

The scientific literature on EMDR supports several conclusions: (1) EMDR is an efficacious treatment for PTSD, (2) the efficacy of EMDR is comparable with that of trauma-focused CBT approaches such as PE, and (3) eye movements and other bilateral stimulation techniques appear to be unnecessary and do not uniquely contribute to clinical outcomes. The characteristic procedural feature of EMDR appears therapeutically inert, and the other aspects of this treatment (e.g., imaginal exposure, cognitive reappraisal, *in vivo* exposure; Shapiro, 2001) overlap substantially with those of exposure-based treatments for PTSD. Despite its status as an evidence-based psychotherapy that is recommended as a first-line treatment for PTSD in clinical guidelines (e.g., Department of Veterans Affairs/Department of Defense, 2010), EMDR offers few, if any, demonstrable advantages over competing evidence-based psychological treatments. Moreover, its theoretical model and purported primary active therapeutic ingredient are not scientifically supported. Accordingly, the scientific status of EMDR characterized by McNally's (1999) maxim, "What is effective in EMDR is not new, and what is new is not effective" (p. 619), still holds today.

In response to findings that bilateral stimulation techniques are

irrelevant to the efficacy of EMDR, Shapiro (2002) emphasized the numerous components that comprise this therapy and claimed that, "as with any complex treatment, the elimination of a single component is likely to have little effect" (p. 8). Such statements risk obfuscating critical empirical and theoretical issues. If bilateral stimulation is unnecessary for EMDR's efficacy, then it is incumbent on EMDR's proponents to specify the essential features of the treatment to permit controlled experiments designed to assess the relative effects of procedural artifacts and the substantive clinical procedure (Grunbaum, 1985). Without a clear specification of the necessary (characteristic) features of treatment, any number of convenient ad hoc accounts can be advanced to explain away disconfirming evidence. Under such conditions, the theory may be difficult or impossible to test unless much more specific predictions are advanced (see Herbert et al., 2000, for a more complete discussion of these issues; see also Lilienfeld, Lynn, & Lohr, Chapter 1, this volume).

Critical Incident Stress Debriefing and Management

Another aspect of the expansion of clinical services in the trauma field has been the development of procedures to prevent PTSD in trauma-exposed individuals. "Psychological debriefing" quickly emerged as a prophylactic approach to minimize the presumed ill effects of occupational exposure to stressful career episodes in first responder groups (Stuhmiller & Dunning, 2000) and rapidly extended its reach to include a broad range of other "exposed" populations. By the time of the 2001 attacks on the World Trade Center, Kadet (2002) reported that as many as 9,000 purveyors of this prophylactic method converged upon New York City, contending that intervention would be needed by essentially anyone even remotely impacted by the atrocities. The most widely promoted of these interventions has been critical incident stress debriefing and management (CISD; Mitchell, 1983, 1988a, 1988b; Mitchell & Everly, 1993, 1995, 1998). CISD is predicated on the assumption that exposure to ostensibly traumatic life events is a sufficient precursor for the development of psychological symptoms that can readily grow to pathological proportions, and that early and proximal intervention involving some element of emotional catharsis is necessary and sufficient to prevent such sequelae.

Early dissemination and promotion of the technique was based on frequent claims of scientific evidence for its efficacy. Yet, no published outcome studies were found despite extensive efforts to uncover them, and, in some cases, data claimed to have formed the basis for key constructs were ultimately conceded to be unavailable (Gist, Woodall, & Magenheimer, 1999). Meanwhile, widespread application had begun to attract serious and independent study. Gist, Lubin, and Redburn (1998) reported studies of debriefing-styled interventions following the crash of a wide-bodied airliner in which 112 passengers died. Their findings from a nearly

complete sample of career firefighters engaged in body recovery and related operations showed no clinically significant impacts on personnel at 2 years postincident, no evidence of superior resolution for debriefed responders versus those who declined, a slight but statistically significant negative trend in resolution indices for those accepting debriefing, and a clear preference for informal sources of support and assistance that correlated strongly with effective resolution. Other studies have replicated and amplified these core conclusions (Bisson, Jenkins, Alexander, & Bannister, 1997; Carlier, Lamberts, van Uchelen, & Gersons, 1998; Deahl, Gillham, Thomas, Dearle, & Strinivasan, 1994; Griffiths & Watts, 1992; Hobbs, Mayou, Harrison, & Worlock, 1996; Kenardy et al., 1996; Lee, Slade, & Lygo, 1996; McNally, Ehlers, & Bryant, 2003; Rose & Bisson, 1999).

A meta-analysis by Van Emmerick, Kampuis, Hulsbosch, and Emmelkamp (2002) of seven outcome studies revealed that the effect size for CISD was not different than zero (including a range of negative values within its 95% confidence interval) and that CISD was less effective than either nonintervention control or alternative interventions. A widening range of guidelines for evidence-based practice, including the Cochrane Reviews (Rose, Bisson, & Wesselly, 2004), the UK National Institute for Clinical Excellence (NICE, 2005), the World Health Organization (2006), and the Australian Centre for Posttraumatic Mental Health (2007), have offered specific recommendations contraindicating routine debriefing in the wake of trauma.

Although these findings have progressively reduced the endorsement of the technique in many circles, affinity and adherence among many of its core advocates have paradoxically intensified. Indeed, CISD subsequently was modified and transformed into critical incident stress management (CISM), though little in the way of substantive difference in form or application was established (Devilly, Gist, & Cotton, 2006). Gist et al. (1999) suggested that the intensification of promotion is an essential feature of the "groupthink" processes that commonly fuel and maintain pseudoscientific enterprises, and of the "true believer" characteristics of those attracted to such movements.

The International Critical Incident Stress Foundation (2012) was founded by Jeffrey Mitchell and George Everly to promote CISD and CISM, and to provide a vehicle for marketing the training and dissemination of promotional literature. Internal Revenue Service returns for the tax-exempt foundation reveal that its revenues have fallen sharply (by nearly half in the past 5 years); workshop frequency and attendance have dwindled; staff have been jettisoned; and salaries of remaining positions have been reduced. Meanwhile, the number of "CISM teams" maintaining certification through the organization has continued a steep downward trend.

Having built its initial following from the volunteer fire service, the circle came to full closure when the National Fire Protection Association (2012) voted overwhelmingly to remove reference to CISM from its Standard on Fire Department Occupational Safety and Health Program. Other

core proponents, however, had already opted for the "reinvention" tactic. CISD has been recast as "psychological first aid" for several years (e.g., Everly & Flynn, 2006; Parker, Everly, Barnett, & Links, 2006). Despite the efforts of the National Child Traumatic Stress Network to brand and market Psychological First Aid (PFA) as a proprietary program, the majority of references to PFA cited in the PubMed database come from CISD promoters and the *International Journal of Emergency Mental Health*, published by Chevron Publishing, a subsidiary of ICISF.

SCIENCE-BASED TREATMENTS FOR TRAUMA-RELATED DISORDERS

Cognitive-Behavioral Theory

Although numerous psychosocial theories have been advanced to explain the etiology of PTSD, behavioral and cognitive models have the strongest empirical support and are the most widely accepted by the scientific community (Foa & Meadows, 1997; Hofmann & Smits, 2008; Tolin, 2010). The behavioral model of PTSD, as originally posited by Keane, Zimering, and Caddell (1985), suggests that traumatized individuals acquire conditioned fears to a wide assortment of trauma-related stimuli and subsequently avoid those stimuli. Through the processes of higher-order conditioning and generalization, the number of feared stimuli continues to increase even after the trauma has occurred. The cognitive model later introduced by Foa, Steketee, and Rothbaum (1989) holds that PTSD develops when the traumatic event reinforces negative beliefs concerning one's safety and competence. Individuals with PTSD often believe the world to be dangerous and unsafe and therefore live in almost constant fear. They may also believe themselves to be incompetent in effecting control, and they are reluctant to confront challenging situations. Other cognitive models focus on such basic processes as attention, memory, and probability estimation (e.g., Litz & Keane, 1989). According to these models, PTSD develops and/or is maintained when trauma- or threat-related information receives preferential information over less threatening information. This processing bias leads to distorted ways of perceiving and understanding the world.

Cognitive-Behavioral Treatments

Cognitive and behavioral models of PTSD have informed several treatment strategies. We describe those with a substantial body of scientific evidence to address treatment efficacy, and we limit our literature review to RCTs. These are efficacy studies in which people diagnosed with PTSD have been randomly assigned to treatment conditions, at least one of which is designed to control for possible artifacts, such as placebo effects and statistical regression. Although many different treatments fall under the rubric

of "cognitive-behavioral," we limit our discussion to three broad classes of interventions for PTSD: prolonged exposure (PE), anxiety management training (AMT), and cognitive processing therapy (CPT).

Prolonged Exposure

Exposure-based treatments are predicated on the notion that exposures to feared stimuli facilitate habituation or extinction of learned fear (Tryon, 2005). Imaginal exposure involves instructions to imagine the traumatic event as vividly and as fully as possible. For example, a combat veteran could be instructed to imagine his experiences in detail, including visual, auditory, and olfactory features of the event. *In vivo* exposure generally involves the construction of a hierarchy of safe but avoided exposures, with encouragement to confront them gradually. For example, a rape victim who is afraid to stay home alone could be instructed to remain at home for progressively longer periods of time. Such terms as "prolonged exposure" or "flooding" are often used to describe imaginal and/or *in vivo* exposure exercises in which exposure is applied for prolonged periods of time. Systematic desensitization includes imaginal exposure, which is paired with an anxiety-incompatible response, such as relaxation. Typically, the exposures in systematic desensitization are more gradually applied, and for briefer periods of time, than in PE.

An early randomized control trial of prolonged exposure for PTSD examined imaginal exposure with combat veterans (Keane, Fairbank, Caddell, & Zimering, 1989). Compared with a wait-list control condition, PE produced greater reductions in PTSD symptom severity on both standardized measures and clinician ratings. Treatment gains were maintained at 6-month follow-up. Similar results were reported by Brom, Kleber, and Defares (1989) using systematic desensitization for civilians who had experienced a wide variety of traumatic experiences. A treatment combining imaginal and *in vivo* exposure was superior to supportive counseling and wait list, and marginally superior to stress inoculation training (see next section) for sexual-assault-related PTSD (Foa et al., 1999; Foa, Rothbaum, Riggs, & Murdock, 1991). In a mixed sample of civilians with PTSD, Marks, Lovell, Noshirvani, and Thrasher (1998) showed that imaginal PE was roughly equivalent to cognitive therapy (CT; Tarrier et al., 1999a). Combined imaginal and *in vivo* PE was shown to be superior to wait list, comparable to CT, and superior to progressive muscle relaxation training (PMR). The PE component is no less effective than a treatment that combines PE with cognitive restructuring (Foa et al., 2005). The effectiveness of PE was examined by Aderka, Gillihan, McLean, and Foa (2013) in a reanalysis of the data from Foa et al. (2005). Multilevel mediational analyses showed that PE primarily affects PTSD symptoms, which in turn reduce depressive symptoms. ASD can also be treated efficaciously with CBT that contains a PE component (Bryant, Harvey, Dang, Sackville, & Basten, 1998; Bryant, Moulds, Guthrie, & Nixon, 2005; Bryant, Moulds,

& Nixon, 2003; Bryant, Sackville, Dang, Moulds, & Guthrie, 1999). Moreover, there is no empirical evidence that PE presents greater risk for harm than any treatment (Foa et al., 2013). The development and use of exposure-based therapies to treat PTSD and related anxiety disorders is one of the great success stories of the mental health field (Frueh, Turner, & Beidel, 1995; Olatunji, Deacon, & Abramowitz, 2009; Tryon, 2005). Such therapies now constitute the first-line treatments for such disorders, although their degree of dissemination falls far short of the evidential warrant of their effectiveness (Foa et al., 2013).

Anxiety Management Training

AMT, also known as stress inoculation training, refers to an armamentarium of cognitive and behavioral strategies designed to reduce symptoms of anxiety, irritability, and hyperarousal. These techniques include psychoeducation, which involves teaching patients about trauma and typical responses to it; relaxation training, in which patients are taught to reduce muscle tension; breathing retraining, which aims to prevent hyperventilation; self-instruction, in which patients are taught to solve problems by "coaching" themselves mentally; communication training, which is aimed at improving interpersonal functioning; and cognitive therapy, which is aimed at modifying dysfunctional expectations of danger, beliefs about the inability to control stressors, and other errors in interpretations regarding the nature and occurrence of stressful events. Such protocols have been compared with PE in two RCTs. Using a civilian sample of assault victims, Foa et al. (1991) showed that AMT reduced symptoms of PTSD, although the effects were slightly less pronounced than those produced by combined and imaginal reexposure. In a later study, Foa et al. (1999) later compared PE, AMT, and the combination of PE and AMT with a wait-list control condition among assault victims. AMT decreased symptoms using standardized measures of PTSD, although exposure produced greater effects. There was no significant difference between AMT and the combined treatment, and the combined treatment was less effective than PE alone. Thus, AMT by itself appears to be an effective treatment for PTSD.

Cognitive Processing Therapy

Alternative treatment procedures deemphasize the behavioral (task-specific) content of treatment and instead focus on the interpretational, attitudinal, expectational, and memory-based means of modifying stress-related symptoms. Although such procedures are usually targeted at depressive symptoms, they have been applied to PTSD (Marks et al., 1998). Tarrier et al. (1999a) compared cognitive therapy (CT) with imaginal PE for mixed civilian trauma and showed that both were superior to a wait list and were comparable in symptom reduction at the end of treatment at both 6-month and 12-month follow-ups (Tarrier et al., 1999b). A 5-year follow-up of 54% of

the original treatment completers found that participants in the CT condition reported fewer PTSD symptoms than those in the imaginal PE condition (Tarrier & Sommerfield, 2004).

Another cognitive-behavioral application has emphasized the "processing" of traumatic memories as the central feature of treatment. The CPT model was developed for those individuals, primarily women, who had experienced one or more sexual assaults that preceded the development of PTSD (Resick & Schnicke, 1992). Treatment focuses on modifying the dysfunctional beliefs that develop following sexual assault. For example, postassault dysfunctional beliefs regarding the generality of personal danger are challenged Socratically in terms of their "truth value." Behavioral tests are used to determine their accuracy and prospective applicability. For example, if a postassault patient believes she is incompetent to avoid danger and harm, suggestions for the assumption of personal agency or responsibility may be posed as a means of testing the veracity of the belief. Should she act in a competent manner, that fact is used as evidence that personal competence is greater than previously assumed.

The theory underlying CPT proposes that the experience of sexual assault influences people's cognitive "schemata" about themselves, other people, and the world (Horowitz, 1976). More specifically, the rape experience influences the victim's sense of safety, trust, power, competence, esteem, and intimacy in ways that have a negative impact on schemata (McCann, Sakheim, & Abramson, 1988). CPT is focused on identifying such maladaptive schemata and their personal significance in more adaptive processing of rape-related beliefs and attributions.

The theory of PTSD maintenance prescribes that the cognitive content of CPT is the specific and characteristic feature. The victim is first provided with psychoeducational information about the functional connection between cognitive processes as antecedents of negative emotions and maladaptive behavior. The process involves the consideration of cognitive schemata about self, others, and the world and the way in which those schemata relate to safety, trust, power, competence, esteem, and intimacy. This goal is accomplished with the use of narrative impact statements that the client writes at the beginning and the end of treatment. During treatment, the therapist uses Socratic questioning to call into challenge the "truth value" of maladaptive thinking patterns about the personal significance of the assault experience.

The process of treatment also involves more specific narrative recounting to the worst-case assault experience as the focus of two early sessions of treatment in which negative emotions are connected to the event. Clients are also asked to read the narrative each day before the subsequent treatment session. The narrative exposure component is intended to activate the maladaptive schemata and allow corrective information to be processed into more adaptive schemata. The treatment process is prescriptive, with specific treatment activities scheduled for each of the 12 sessions (Resick & Schnicke, 1993).

The efficacy of CPT has been addressed in a number of outcome studies. Resick and Schnicke (1992) randomly assigned rape victims to either a wait list or CPT treatment condition. Treatment was administered in group format for 12 sessions. Analysis of pre–post PTSD and depression measures showed statistically significant reductions in the CPT group. The difference in improvement was maintained at 3- and 6-month follow-up. In the CPT group, there was also a reduction in depressive symptoms from posttreatment to 3-month follow-up. At 3 months posttreatment, only 2 of 19 women in the CPT condition met the criteria for PTSD, and none met PTSD criteria at 6-month follow-up.

A second study compared CPT and PE with a minimal attention (MA) control condition (Resick, Nishith, Weaver, Astin, & Feuer, 2002). Women in the MA condition were informed that they would be provided treatment 6 weeks following administration of assessment procedures for all participants. They were told that during the interim they could call and talk to a therapist should the need arise. Analysis of PTSD, depression, and guilt measures showed statistically significant reductions in both treatment conditions relative to the MA condition. Direct comparisons between PE and CPT in those completing treatment revealed small to moderate effects sizes in favor of CPT on measures of PTSD and depression at posttreatment and 3-month follow-up. At 9-month follow-up, a small effect size favored PE. In a reanalysis of the data, Gallagher and Resick (2012) found that CPT reduced hopelessness, and this predicted reductions in PTSD symptoms. PE also reduced PTSD symptoms but did so independently of changes in hopelessness. Thus, a reduction of hopelessness may be a major mechanism of action for CPT.

A number of studies derived from the findings of Resick et al. (2002) have addressed issues of treatment effectiveness. A 5-year follow-up study on 87% of treatment completers showed that reductions in symptoms were maintained with no statistically significant differences between treatment conditions. Seventy-eight percent of those who participated in CPT no longer met the criteria for PTSD, and 83% of those who participated in PE no longer met the criteria for PTSD.

Nishith, Resick, and Griffin (2002) examined the pattern of clinical change over the course of the treatments. Reexperiencing symptoms increased in both groups before improvement occurred. Hyperarousal symptoms remained constant until the fourth treatment session and gradually declined. Avoidance symptoms and total symptoms also declined after the fourth session. Participants in CPT showed a linear decline and participants in PE showed a quadratic decline, indicating that intensive PE first produced an increase in avoidance symptoms that did not occur in CPT. Rizvi, Vogt, and Resick (2009) examined the predictive value of personal characteristics in the reduction of trauma-related symptoms. The dropout rate was greater in younger, less intelligent, and less educated participants. Depression, trait anger, and guilt did not predict dropout in general, but trait anger predicted greater dropout in the PE condition. Treatment

efficacy was unrelated to age, intelligence, and education, but younger women displayed better outcome with CPT, older women in PE had the next best outcome, and younger women in PE had the poorest outcomes. Greater depression and guilt predicted greater reductions in PTSD symptoms in both treatment conditions.

Nishith, Nixon, and Resick (2005) examined the potential effects of comorbidity between major depression and PTSD in the treatment of CPT for guilt. Participants comorbid for PTSD and depression showed higher depression and PTSD symptoms at intake relative to participants without depression. There was no difference in effectiveness in the comorbid participants relative to the PTSD-only participants. However, CPT showed greater effectiveness than PE. As sleep-related problems and health-related problems are common in PTSD (Green & Kimerling, 2004; Ohayon & Shapiro, 2000), Galovski, Monson, Bruce, and Resick (2009) examined changes in these variables following CPT and PE. The results showed that sleep quality was improved by both treatments and at 9-month follow-up. Health-related concerns were also reduced in both treatments but more so in CPT.

Demonstration of the efficacy and effectiveness of a CPT protocol should be followed by attempts to identify the specific or characteristic features of the treatment that are primarily responsible for the beneficial effects (Borkovec & Castonguay, 1998; Lohr et al., 2003). Resick et al. (2008) conducted such a study to examine the characteristic features of CPT. Participants were 150 women diagnosed with PTSD who had an extensive history of interpersonal violence. Participants were randomly assigned to one of three treatment groups: complete CPT; CPT-C, which was identical to CPT but without the written, narrative trauma accounts; and only written accounts (WA). Treatment was administered in group format in 7 to 12 sessions across a period of 6 weeks, during which treatment adherence and competence were independently assessed. Standardized measures of PTSD and depression were administered pretreatment, weekly during treatment, posttreatment, and at 6-month follow-up. Participants in all three conditions improved on measures of PTSD and depression, with no statistically significant differences among the treatment conditions. Thus, it appears that a form of narrative recounting is an active component of treatment, as is the component specific to cognitive processing. However, the findings suggest there is little beneficial interaction of the two components. In sum, PE, AMT, and CPT all appear to be efficacious, and CPT may possess greater effectiveness for related clinical conditions. In a reanalysis of Resick et al. (2008), Liverant, Suvak, Pineles, and Resick (2012) further examined treatment effectiveness. Multilevel regression analyses showed that the reductions of PTSD and depressive symptoms were strongly related and that they were reduced independently and concurrently. Thus, CPT appears to have a highly desirable side effect for depressive impairments that are common in sexual assault-related PTSD.

Optimism about the efficacy of CPT, however, must be cautious. There are only three published studies on CPT efficacy, whereas PE and AMT have been subjected to far more empirical tests. Moreover, the investigators who developed CPT are those who published the studies demonstrating efficacy, raising the possibility of researcher allegiance effects (Gaffan, Tsaousis, & Kemp-Wheeler, 1995; Leykin & DeRubeis, 2009; Luborsky et al., 1999; Tolin, 2010).

FUTURE DIRECTIONS

Psychological First Aid

The failure of CISD/CISM to achieve its stated objectives and, more pointedly, its tendency to impede coping reactions among particularly sensitive subgroups created an interesting conceptual and practical vacuum. Although resilience is the modal trajectory in the face of trauma (Bonanno, 2004), it has become a social expectation that "mental health assistance" will follow any untoward experience in life, be it in the workplace, the schools, or the community. "Critical incident" programs are an essential feature of workplace employee assistance programs, school crisis response plans, and community response programs such as the U.S. Department of Homeland Security's Community Emergency Response Teams (CERT) and the U.S. Department of Health and Human Services' Medical Reserve Corps. It was all but inevitable that some program, hopefully one with a more substantial evidence base, would appear to fill the interventionist void.

The most promising candidate was the psychological first aid (PFA) project of the National Child Traumatic Stress Network (NCTSN), carried out in partnership with the Veterans Administration's National Center for PTSD (NCPTSD). This project was the work of an eight-member core team, drawing on their academic resources and practical experience to create an "evidence-informed" intervention (Brymer et al., 2006) in the form of a "Field Operations Guide." Included are "eight core actions" that PFA implements:

1. Contact and engage with affected persons in a nonintrusive, compassionate, and helpful manner.
2. Provide safety and comfort to enhance immediate coping.
3. Stabilize emotionally overwhelmed and distraught survivors.
4. Gather information to identify immediate needs and concerns.
5. Offer practical help in meeting needs and concerns.
6. Connect with social supports, including family, friends, and community helping resources.
7. Obtain information to aid in coping with the event and aftermath.
8. Link survivors with immediate and likely needed services.

Since that time, PFA has been expanded to include versions for chaplains and pastors, CERT groups, Medical Reserve Corps volunteers, schools, and other constituencies; many workshops have been held; web-based training has been developed; and Smartphone applications are under development.

Despite the outpouring of resources and effort that has gone into promoting the product, a void exists with respect to its evaluation. Reports have appeared in various service venues examining the perceived applicability of the techniques and providing Likert-style feedback regarding perceptions of providers trained to administer PFA (e.g., Allen et al., 2010). Moreover, descriptions of the material proclaiming its applicability and promoting its further dissemination abound (Brymer, Steinberg, Sornborger, Layne, & Pynoos, 2008; Ruzek et al., 2007; Vernberg et al., 2008). Although the advocates of PFA indicate that outcome research is needed, little has been accomplished.

These and other considerations suggest that the promotion of PFA may be replicating rather than rectifying the pseudoscientific practices of CISD. These practices include:

1. *Limited claims versus expansive reach.* Promoters of CISD claimed that it was preventative for any population in essentially any situation. Initially targeted to a limited population of first responders in a limited set of circumstances, it was quickly taught to all comers and applied to numerous situations from workplace to schools, from terrorism to corporate downsizing. PFA was initially targeted toward mental health personnel assisting victims of disaster, but is now promoted for essentially the same, nearly unlimited expanse. Neither program, however, has provided evidence of applicability before venturing into expanded areas of application, nor has either adduced evidence of impact across these applications and settings.

2. *Measurement of outcomes versus indices of receptivity.* Both CISD and PFA have reported subjective responses of recipients as if they were indicators of outcome (Allen et al., 2010; Robinson & Mitchell, 1993). Subjective reports of benefits may result in the belief that the activity is helpful and that the training aided in the delivery of the intervention. However, these beliefs do not ensure that the intervention delivered actually made a difference in a defined and measurable outcome. Indeed, CISD data demonstrated that anecdotal endorsements show no systematic relationship to clinical outcome (Devilly et al., 2006). The absence of any reported data regarding actual outcomes for PFA raises similar concerns.

3. *Portrayal of the ubiquitous as if uniquely innovative.* CISD was the packaging of the basic social psychological group process for market to a specific service provision niche. PFA may also represent the recitation of basic common-sense palliative actions, known for generations of informal application as a branded psychological intervention.

4. *Overstatement of evidential basis.* CISD claimed to be a "proven"

approach with wide-ranging evidence to support efficacy and utility when little actual evidence existed. Over a decade of substantive scientific inquiry, evidence accumulated by independent researchers not only failed to support those assertions but contradicted them (Devilly et al., 2006). PFA, though declaring itself to be an "evidence-informed" approach, has yet to provide any mapping of its purported evidence base onto its specified actions and has adduced no data in support of its efficacy or effectiveness. The evidence base represents, at best, the consensus of a select group of authors regarding best practices. This type of evidence occupies the lowest tier in most evidentiary hierarchies accepted for evidence-based practice (National Institute for Clinical Excellence, 2004, p. 7-6; Oxford Centre for Evidence Based Medicine, 2012).

The notion of psychological first aid is hardly new or unique. It dates back at least 65 years (Blain, Hoch, & Ryan, 1945), and many independent variations exist today (e.g., Burke & Richardson, 2012; Everly & Flynn, 2006; Kitchener & Jorm, 2002; World Health Organization, 2011). Their commonality is that they all advance a collection of pragmatic folk wisdom and basic helping steps with little scientific grounding. The promotion of named or "branded" applications, as if they were "evidence informed," is questionable. Defending them as if they represented a proprietary product or patent remedy, while they lack compelling empirical support, suggests the promotion of pseudoscientific techniques.

Resilience Training for Postevent High-Risk, High-Impact Groups

Conceptual Issues

To understand variations among individuals in how they respond to stressful events, researchers have examined risk factors for pathological reactions, typically using post hoc designs. Such studies tell us something about the people who have had maladaptive reactions to life's adversities—their personality profiles, life experiences and histories, and usual methods of coping with adversity—all through retrospective lenses. Such approaches are informative regarding the nature and course of pathological reactions, but they tell us little about why people stay healthy.

Three distinct but related questions arise:

1. Why do some people who experience severe life trials bounce back to normal functioning very quickly?
2. What helps people to resolve traumatic experiences over time?
3. What information is needed to inoculate people most at risk (e.g., emergency or military service) against unhealthy reactions?

The last question has recently led organizational and clinical consultants to offer resilience training as a means of promoting health under

severe risk. However, there is scant evidence for the use of any form of resilience, and there is even less agreement regarding what resilience means.

Three groups of resilient outcomes have been distinguished: those in which (1) at-risk individuals show better than expected outcomes; (2) individuals maintain positive adaptation despite stressful experiences; and (3) individuals show good recovery after a traumatic incident (Masten, 1994; Masten, Best, & Garmezy, 1990). Such a perspective, however, views resilience as an outcome trajectory distinct from the type of responses typically associated with recovery from trauma (Bonnano, 2004). Many individuals who undergo a traumatic event and go on to recover without developing PTSD, often experience subthreshold levels of psychological symptoms and significant disruption of daily functioning. In contrast, resilient individuals may experience only mild and transient disruptions in physical and emotional well-being (e.g. sleeplessness, negative affect, difficulty concentrating), with relatively stable levels of adjustment over time (Bonanno, Rennicke, & Dekel, 2005).

Inoculation Training

Bonanno (2005) noted that resilience in adults typically occurs following an isolated and usually brief traumatic event. Under such circumstances, resilience involves immediate and pragmatic forms of coping rather than long-term strategies. To our knowledge, only one published controlled field trial has examined the effects of resilience training on adults following traumatic life events. Nevertheless, it was not an RCT. Sharpley, Fear, Greenberg, Jones, and Wessely (2008) referred to their intervention as predeployment stress briefing when provided to the United Kingdom's armed forces (Royal Navy and Royal Marines) before deployment to the 2003 Iraq War. This intervention consisted of education regarding the "role of the mental health team; an outline of the medical facilities in the Primary Casualty Receiving Facility; definition of stress, pressure and strain; types of stressors (physical, social, occupational, and traumatic); effects of stress on individuals; advice on handling human remains; managing stressful thinking in a chemical or biological environment; simple advice on reducing stress; the importance of morale; levels of support available and when/where to seek this" (Sharpley et al., 2008, p. 31). Upon returning from Iraq, all troops completed a questionnaire regarding their reactions. Those who had received the pre-briefings were designated as the treatment group, and those personnel not registered as having received the pre-briefings were designated as a no-treatment control. As may be expected when using post hoc and self-selected samples, the treatment group differed significantly from the control group on a number of variables. Most notably, they experienced a higher number of traumatic events and more often played a combat role during deployment. The results, though not

statistically significant, pointed toward lowered pathology in the prebriefing group.

In the U.S. military, a plan has recently emerged to develop and use a resilience training program known as comprehensive soldier fitness (CSF; Cornum, Matthews, & Seligman, 2011). Based on concepts drawn from positive psychology (Seligman & Csikszentmihalyi, 2000; Seligman, Steen, Park, & Peterson, 2005), CFS aims to create psychological resilience based on facilitating personal growth, developing skills to reduce the sequelae of traumatic exposure, and helping military personnel and their families cope with the stress of prolonged, dangerous deployments. Proponents of the method claimed on websites (and in varying forms in conference presentations) that CSF

> is a long term strategy that better prepares the Army community—including all Soldiers, Family members, and the Department of the Army Civilian workforce—to not only survive, but also thrive at a cognitive and behavioral level in the face of protracted warfare and everyday challenges of Army life that are common in the 21st Century. The program, based on 30-plus years of scientific study and results, uses individual assessments, virtual training, classroom training, and embedded resilience experts to provide the critical skills our Soldiers, Family members and Army Civilians need. (Comprehensive Soldier Fitness, 2012)

Likewise, a program originally called "Battlemind" and now known as "resilience training" is also being promoted by the U.S. military. Battlemind has been empirically investigated (Adler, Bliese, McGurk, Hoge, & Castro, 2009) but only when used postdeployment, as an alternative to psychological debriefing. Similarly, the Australian Defence Force has been utilizing a program called "BattleSMART." As explained by the national coordinator of the program:

> The resilience training program, dubbed BattleSMART (Self-Management and Resilience Training), is a cognitive-behavioural based program that aims to develop both arousal reduction techniques (i.e., the Self-Management component) and adaptive cognitive coping strategies. (Cohn, Hodson, & Crane, 2010, p. 16)

Despite the good intentions of military leaders, McNally (2012) expressed concern that CFS was implemented without controlled research to determine its efficacy or risks. He noted that it would have been desirable to conduct an RCT to determine whether it reduces rates of postevent PTSD in comparison with the military's standard (previous) program (Tedeschi & McNally, 2011). The paucity of empirical evidence and the methodological limitations of the available evidence have prompted Steenkamp, Nash, and Litz (2013) to question the widespread implementation of CSF and to call for more systematic study of its efficacy and effectiveness.

Pre-event Training for High-Risk, High-Impact Groups

Although no data support the efficacy or effectiveness of the postevent intervention, an alternative approach is to not wait for a high-impact stressor to occur, but to design prevention efforts that could be implemented prior to exposure to a potentially traumatizing event. Varker and Devilly (2012) published am RCT of inoculation (resilience) training using an analogue design. Outcome was based on people's short- and long-term (4 weeks) reactions to watching a stressful video of paramedics attending the scene of a road traffic accident. The study provided a serial approximation to a stressful event, and in the experimental group psychoeducation and coping strategies were used to deal with aversive physiological responses and high levels of stress. In this study, the researchers provided community participants with either inoculation training or pragmatic training dubbed accident management training that provided practical tips and strategies on what to do if participants experienced or witnessed a traffic accident. Both trainings were provided to participants one week before they were exposed to a video that had been used previously to investigate the effects of psychological debriefing (Devilly & Annab, 2008; Devilly & Varker, 2008; Devilly, Varker, Hansen, & Gist, 2007). Those who received the inoculation training fared no worse than the control group on the main outcome measures. However, participants who received the inoculation training displayed improvements in negative affect (especially depression and stress levels), suggesting a more general positive result from the intervention than pragmatic training. These findings give us cause for cautious optimism.

PROFESSIONAL IMPLICATIONS FOR PSEUDOSCIENTIFIC PRACTICES

The American Psychological Association Ethical Principles and Code of Conduct (2002) are intended to hold practitioners accountable to the profession and the scientific discipline (Cottone & Tarvydas, 1998). The practice of pseudoscience in clinical psychology raises ethical implications for the mental health profession. For instance, the code's preamble states that psychologists should work to develop a valid and reliable body of scientific knowledge based on research. Additionally, Principle B (integrity) states that psychologists should seek to promote integrity in the science, teaching, and practice of psychology. The General Standard, Basis for Scientific and Professional Judgments (1.06), states that psychologists should rely on scientifically and professionally derived knowledge when making scientific or professional judgments or when engaging in scholarly or professional endeavors.

There is little question that psychologists have the primary ethical obligation of *Primum non nocere*: First Do No Harm (either wittingly or unwittingly). This obligation is most directly applied to acts of commission for which the public will incur significant cost, damage, or both.

Therefore, psychologists have an ethical obligation to be informed of clinical procedures that have been shown to be potentially harmful (Barlow, 2010a; Lilienfeld, 2007). The obligation, however, also applies to clinical acts of commission in which there is evidence that the purported benefits of a treatment are absent, or in which there is an absence of evidence that the treatment is beneficial. In the latter case, the application of such a treatment can be ethically defensible if the clinical psychologist is aware that little evidence exists, and when the clinical psychologist informs the client of that fact. That is, the client should be informed (1) of the fact that the treatment is not "usual and customary," (2) of what the usual and customary treatments are, and (3) of the rationale for applying the treatment in addition to, or instead of, the usual and customary treatments. In this way, the client can make an informed decision about participating in an "experimental" treatment application. In addition, the psychologist can avoid the professional pitfalls of overpromoting a treatment to which he or she has an allegiance (Luborsky et al., 1999).

Where there exists evidence that a treatment is ineffective or that a treatment's effectiveness is due to factors other than the specificity of the treatment, ethical obligations also apply. As in the case of EMDR, the clinical practitioner has two obligations: inform the client (1) of alternative, efficacious treatments, and (2) inform the client that the specific features of the treatments (eye movements) are inert components of the clinical procedure. To provide such informed consent, the practicing clinician must keep abreast of the research on the efficacy and effectiveness of those treatments. To address the problem of pseudoscience in the mental health domain, practitioners must actively embrace their professional and ethical obligation to be knowledgeable about the empirical basis for their treatment. This mandate compels practitioners to stay abreast of the relevant literature, including distinctions between science and pseudoscience (Haas & Malouf, 1995).

The profession of psychology requires specialized skills and methods that are based on scientific knowledge (Sinclair, Simon, & Petifor, 1996). However, if there is a proliferation of treatments that are not based on scientific principles, the implications for the status of professional clinical psychology are substantial. Mental health practitioners may be increasingly vulnerable to pseudoscientific promotion, given that accelerating demand for clinical services has sometimes led to reduction of scientific training (Dawes, 1994; Singer & Lalich, 1996). By appealing to more explicit criteria for methodological precision in treatment validation (e.g., Borkovec & Castonguay, 1998; Foa & Meadows, 1997), practitioners may begin to ask specific questions aimed at determining the efficacy of a given treatment. These questions should include the following:

Does the treatment work?
If the treatment works, how does it work?
What are the mechanisms of change?

Questions can then be addressed regarding the theoretical foundations of the treatment. Such questions regarding theoretical rationale may prove to be vital in separating science from pseudoscience. Practitioners are also encouraged to proceed with caution when presented with treatments that claim to be effective for a wide range of disorders. Scientific interventions typically have clear-cut boundaries in their applicability. Treatments that purport to have limitless boundaries are often indicative of a pseudoscience (see Lilienfeld, Lynn, & Lohr, Chapter 1, this volume).

Mental health practitioners are at serious risk for the acceptance of pseudoscientific trauma services, given that the increased demand for the provision of clinical services may have resulted in a decline of scientific training. It is not surprising, therefore, that the profession is plagued with an industry of procedures lacking in empirical support (Beyerstein, 2001) and are frequently practiced with a trusting laity. The evaluation of all psychosocial treatments must rest on the substantive aspects of procedure rather than on its superficial appearance. The practice and research communities will benefit only when psychosocial interventions are marketed and accepted on the basis of compelling evidence. It is the compelling evidence, and the rigor of the research procedures from whence it derives, that will determine whether novel treatments can be incorporated into the corpus of science-based practice.

GLOSSARY

Anxiety management training (AMT): A set of cognitive-behavioral techniques designed to facilitate adaptive coping with stress. AMT components include relaxation training, breathing retraining, psychoeducation, self-instruction, communication training, and cognitive therapy. AMT as actually practiced may contain some or all of these components.

Cognitive-behavioral therapy (CBT): A set of therapeutic techniques based on cognitive and behavioral theories of psychopathology.

Cognitive therapy (CT): A form of psychotherapy that aims to modify dysfunctional beliefs or assumptions. CT techniques include Socratic questioning and behavioral tests.

Critical incident stress debriefing (CISD): A technique used to ward off posttraumatic stress symptoms among trauma-exposed individuals. It is predicated on the assumptions that exposure to ostensibly traumatic life events is a sufficient precursor to psychological symptoms, and that early and proximal intervention involving emotional catharsis is necessary and sufficient to prevent these outcomes.

Exposure: A cognitive-behavioral intervention for anxiety disorders. In exposure therapy, the patient either confronts previously avoided objects or situations *in vivo* or confronts previously avoided thoughts or memories.

Measurement reactivity: The spurious effect of prior administration of an assessment procedure on the results of a second administration of the same assessment procedure. Measurement reactivity may make it appear that an intervening

treatment has resulted in beneficial change when the difference is only an artifact of the measurement process. The possibility of measurement reactivity requires the inclusion of the wait-list control condition to rule out procedural artifacts and reduce the probability of Type I error.

Systematic desensitization: A behavioral treatment procedure that includes the training of relaxation that is paired with imaginal or *in vivo* exposure to feared stimuli in a stepped or graduated manner.

REFERENCES

Aderka, I. M., Gillihan, S. J., McLean, C. P., & Foa, E. B. (2013). The relationship between posttraumatic and depressive symptoms during prolonged exposure with and without cognitive restructuring for the treatment of posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 81*, 1–8.

Adler, A. B., Bliese, P. D., McGurk, S., Hoge, C. W., & Castro, C. A. (2009). Battlemind debriefing and Battlemind training as early interventions with soldiers returning from Iraq: Randomization by platoon. *Journal of Consulting and Clinical Psychology, 77*, 928–940.

Allen, A. B., Brymer, M. J., Steinberg, E. M., Vernberg, E. M., Jacobs, A., Speier, A. H., et al. (2010). Perceptions of psychological first aid among providers responding to Hurricanes Gustav and Ike. *Journal of Traumatic Stress, 23*, 509–513.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2004). *Practice guideline for the treatment of patients with acute stress disorder and posttraumatic stress disorder*. Arlington, VA: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271–285.

Asmundson, G. J., Stapleton, J. A., & Taylor, S. (2004). Are avoidance and numbing distinct PTSD symptoms. *Journal of Traumatic Stress, 17*, 467–475.

Australian Centre for Posttraumatic Mental Health. (2007). *Australian guidelines for the treatment of adults with acute stress disorder and posttraumatic stress disorder*. Melbourne, Victoria, Australia: Author.

Barlow, D. H. (2010a). Negative effects from psychological treatments. *American Psychologist, 65*, 13–20.

Barlow, D. H. (2010b). The dodo bird again and again. *Behavior Therapist*, *33,* 15–16.

Beutler, L. E. (2000). Empirically based decision making in clinical practice. *Prevention and Treatment, 3*, article 27.

Beyerstein, B. L. (2001). Fringe psychotherapies: The public at risk. *Scientific Review of Alternative Medicine, 5*, 70–79.

Bisson, J., & Andrew, M. (2007). Psychological treatment of post-traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, No. CD003388.

Bisson, J. I., Ehlers, A., Matthews, R., Piling, S., Richards, D., & Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder: Systematic review and meta-analysis. *British Journal of Psychology, 190*, 97–104.

Bisson, J. I., Jenkins, P. L., Alexander, J., & Bannister, C. (1997). A randomized controlled trial of psychological debriefing for victims of acute harm. *British Journal of Psychiatry, 171*, 78–81.

Blain, D., Hoch, P., & Ryan, V. G. (1945). A course in psychological first aid and prevention: A preliminary report. *American Journal of Psychiatry, 101*, 629–634.

Bonanno, G. A. (2004). Loss, trauma, and human resilience—Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist, 59*, 20–28.

Bonanno, G. A. (2005). Resilience in the face of potential trauma. *Current Directions in Psychological Science, 14*, 135–138.

Bonanno, G. A., Rennicke, C., & Dekel, S. (2005). Self-enhancement among high-exposure survivors of the September 11th terrorist attack: Resilience or social maladjustment? *Journal of Personality and Social Psychology, 88*, 984–998.

Borkovec, T. D., & Castonguay, L. G. (1998). What is the meaning of "empirically supported therapy? *Journal of Consulting and Clinical Psychology, 66*, 136–142.

Boudewyns, P. A., & Hyer, L. A. (1996). Eye movement desensitization and reprocessing (EMDR) as treatment for post-traumatic stress disorder. *Clinical Psychology and Psychotherapy, 3*, 185–195.

Boudewyns, P. A., Stwertka, S. A., Hyer, L. A., Albrecht, J. W., & Sperr, E. V. (1993). Eye movement desensitization for PTSD of combat: A treatment outcome pilot study. *Behavior Therapist, 16*, 29–33.

Bradley, R., Green, J., Russ, E., Dutra, L., & Westen, D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *American Journal of Psychiatry, 162*, 214–227.

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community. *Archives of General Psychiatry, 55*, 626–632.

Brom, D., Kleber, R. J., & Defares, P. B. (1989). Brief psychotherapy for posttraumatic stress disorders. *Journal of Consulting and Clinical Psychology, 57,* 607–612.

Bryant, R. A., Harvey, A. G., Dang, S. T., Sackville, T., & Basten, C. (1998). Treatment of acute stress disorder: A comparison of cognitive-behavior therapy and supportive counseling. *Journal of Consulting and Clinical Psychology, 66*, 862–866.

Bryant, R. A., Moulds, M. L., Guthrie, R. M., & Nixon, R. D. V. (2005). The additive benefit of hypnosis and cognitive-behavior therapy in treating acute stress disorder. *Journal of Consulting and Clinical Psychology, 73*, 334–340.

Bryant, R. A., Moulds, M. L., & Nixon, R. D. V. (2003). Cognitive-behavior therapy of acute stress disorder: A four-year follow-up. *Behaviour Research and Therapy, 41*, 489–494.

Bryant, R. A., Sackville, T., Dang, S. T., Moulds, M., & Guthrie, R. (1999). Treating acute stress disorder: An evaluation of cognitive-behavior therapy and supportive counseling. *American Journal of Psychiatry, 156*, 1780–1786.

Brymer, M., Jacobs, A., Layne, C., Pynoos, R., Ruzek, J., Steinberg, A., et al. (2006). *Psychological first aid: Field operations guide* (2nd. ed.). National Center for PTSD. Retrieved February 1, 2009, from *www.ncptsd.va.gov/ncmain/ncdocs/manuals/PFA_2ndEditionwithappendices.pdf*.

Brymer, M. J., Steinberg, A. M., Sornborger, J., Layne, C. M., & Pynoos, R. S. (2008). Acute interventions for refugee children and families. *Child and Adolescent Psychiatric Clinics of North America, 17*, 625–640.

Burke, S., & Richardson, J. (2012). *Psychological first aid: An Australian guide*. Melbourne: Australian Psychological Society.

Cahill, S. P., Carrigan, M. H., & Frueh, B. C. (1999). Does EMDR work? If so, Why?: A critical review of controlled outcome and dismantling research. *Journal of Anxiety Disorders, 13*, 5–33.

Carlier, I. V. E., Lamberts, R. G., van Uchelen, A. J., & Gersons, B. P. R. (1998). Disaster related post-traumatic stress in police officers: A field study of the impact of debriefing. *Stress Medicine, 14*, 143–148.

Carrigan, M. H., & Levis, D. J. (1999). The contributions of eye movements to the efficacy of brief exposure treatment for reducing fear of public speaking. *Journal of Anxiety Disorders, 13*, 101–118.

Cohn, A., Hodson, S., & Crane, M. (2010, April 16–17). Resilience training in the Australian Defence Force. *InPsych, 8*.

Comprehensive Soldier Fitness. (2012). *Frequently asked questions*. Retrieved August 31, 2012, from *http://csf.army.mil/faq.html#*.

Cornum, R., Matthews, M. D., & Seligman, M. E .P. (2011). Comprehensive soldier fitness: Building resilience in a challenging institutional environment. *American Psychologist, 66*, 4–9.

Cottone, R. R., & Tarvydas, V. M. (1998). *Ethical and professional issues in counseling*. Upper Saddle River, NJ: Prentice-Hall.

Cusak, K., & Spates, C. R. (1999). The cognitive dismantling of eye movement desensitization and reprocessing (EMDR) treatment of posttraumatic stress disorder (PTSD). *Journal of Anxiety Disorders, 13*, 87–99.

David, D., & Montgomery, G. H. (2011). The scientific status of psychotherapies: A new evaluative framework for evidence-based psychosocial intervention. *Clinical Psychology: Science and Practice, 18*, 89–99.

Davidson, P. D., & Parker, C. H. (2001). Eye movement desensitization and reprocessing (EMDR): A meta-analysis. *Journal of Consulting and Clinical Psychology, 69*, 305–316.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.

Deahl, M. P., Gillham, A. B., Thomas, J., Dearle, M. M., & Strinivasan, M. (1994). Psychological sequelae following the Gulf War: Factors associated with subsequent morbidity and the effectiveness of psychological debriefing. *British Journal of Psychiatry, 165*, 60–65.

DeBell, C., & Jones, R. D. (1997). As good as it seems?: A review of EMDR experimental research. *Professional Psychology: Research and Practice, 28*, 153–163.

Department of Veterans Affairs/Department of Defense. (2010). *VA/DoD clinical*

practice guideline for management of post-traumatic stress. Washington, DC: Author.

Devilly, G. J. (2001a). The roles of popularised distraction during exposure and researcher allegiance during outcome trails. *Behavior Therapist*, *24*, 18–21.

Devilly, G. J. (2001b). Effect size and methodological rigor in EMDR: A reply to Lipke's comment. *Behavior Therapist, 24*, 195–196.

Devilly, G. J. (2002). Eye movement desensitization and reprocessing: A chronology of its development and scientific standing. *Scientific Review of Mental Health Practice, 1*, 113–138.

Devilly, G. J., & Annab, R. (2008). A randomised controlled trial of group debriefing. *Journal of Behavior Therapy and Experimental Psychiatry, 39*, 42–56.

Devilly, G. J., Gist, R., & Cotton, P. (2006). Ready. Fire. Aim!: The evolution of psychological debriefing services and intervention outcome. *Review of General Psychology, 10*, 318–345.

Devilly, G. J., & Lohr, J. M. (2008). Science and pseudoscience in victims services. *International Journal of Victimology, 15*, 105–122.

Devilly, G. J., & Spence, S. H. (1999). The relative efficacy and treatment distress of EMDR and a cognitive-behavior trauma treatment protocol in the amelioration of posttraumatic stress disorder. *Journal of Anxiety Disorders, 13*, 131–157.

Devilly, G. J., Spence, S., & Rapee, R. (1998). Statistical and reliable change with eye movement desensitization and reprocessing: Treating trauma within a veteran population. *Behavior Therapy, 26*, 435–455.

Devilly, G. J., & Varker, T. (2008). The effect of stressor severity on outcome following group debriefing. *Behaviour Research and Therapy, 46*, 130–136.

Devilly, G. J., Varker, T., Hansen, K., & Gist, R. (2007). The effects of psychological debriefing on eyewitness memory. *Behaviour Research and Therapy, 45*, 1245–1254.

Dunn, T. M., Schwartz, H., Hatfield, R. W., & Wiegele, M. (1996). Measuring effectiveness of eye movement desensitization and reprocessing (EMDR) in non-clinical anxiety: A multi-subject, yoked control design. *Journal of Behavior Therapy and Experimental Psychiatry, 27*, 231–239.

Everly, G. S., Jr., & Flynn, B. W. (2006). Principles and practical procedures for acute psychological first aid training for personnel without mental health experience. *International Journal of Emergency Mental Health, 8*, 93–100.

Feske, U., & Goldstein, A. J. (1997). Eye movement desensitization and reprocessing treatment for panic disorder: A controlled outcome and partial dismantling study. *Journal of Consulting and Clinical Psychology, 65*, 1026–1035.

Foa, E. B., Dancu, C. V., Hembree, E. A., Jaycox, L. H., Meadows, E. A., & Street, G. P. (1999). A comparison of exposure therapy, stress inoculation training and their combination for reducing PTSD in female assault victims. *Journal of Consulting and Clinical Psychology, 67*, 194–200.

Foa, E. B., Gillihan, S. J., & Bryant, R. A. (2013). Challenges and successes in the dissemination of evidence-based treatments for posttraumatic stress: Lessons learned from prolonged exposure therapy for PTSD. *Psychological Science in the Public Interest, 14*, 65–111.

Foa, E. B., Hembree, E. A., Cahill, S. P., Rauch, S. A. M., Riggs, D. S., Feeney, N. C., et al. (2005). Randomized trial of prolonged exposure for posttraumatic stress disorder with and without cognitive restructuring: Outcome at

academic and community clinics. *Journal of Consulting and Clinical Psychology, 73*, 953–964.
Foa, E. B., & Meadows, E. A. (1997). Psychosocial treatments for posttraumatic stress disorder: A critical review. *Annual Review of Psychology, 48*, 449–480.
Foa, E. B., Rothbaum, B. O., Riggs, D. S., & Murdock, T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology, 59*, 715–723.
Foa, E. B., Steketee, G., & Rothbaum, B. O. (1989). Behavioral/cognitive conceptualizations of post-traumatic stress disorder. *Behavior Therapy, 20*, 155–176.
Foley, T., & Spates, C. (1995). Eye movement desensitization of public speaking anxiety: A partial dismantling study. *Journal of Behavior Therapy and Experimental Psychiatry, 26*, 321–329.
Fox, R. E. (2000, Winter). The dark side of evidence-based treatment. *Practitioner Focus*. Available at *www.apa.org/practice/pf/jan00/cappchair.html*.
Frueh, B. C., Turner, S., & Beidel, D. (1995). Exposure therapy for combat-related PTSD: A critical review. *Clinical Psychology Review, 15*, 799–817.
Gaffan, E. A., Tsaousis, I., & Kemp-Wheeler, S. M. (1995). Researcher alliance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 63*, 966–980.
Gallagher, M. W., & Resick, P. A. (2012). Mechanisms of change in cognitive processing therapy and prolonged exposure therapy for PTSD: Preliminary evidence for the differential effects of hopelessness and habituation. *Cognitive Therapy and Research, 36*, 750–755.
Galovski, T. E., Monson, C., Bruce, S. E., & Resick, P. A. (2009). Does cognitive-behavioral therapy for PTSD improve perceived health and sleep impairment? *Journal of Traumatic Stress, 22*, 197–204.
Gambrill, E. (2012). *Propaganda in the helping professions*. New York: Oxford University Press.
Gaudiano, B. A., & Herbert, J. D. (2000). Can we really tap our problems away?: A critical analysis of thought field therapy. *Skeptical Inquirer, 24*, 29–33, 36.
Gist, R., Lubin, B., & Redburn, B. G. (1998). Psychosocial, ecological, and community perspectives on disaster response. *Journal of Personal and Interpersonal Loss, 3*, 25–51.
Gist, R., Woodall, S. J., & Magenheimer, L. K. (1999). And then you do the Hokey-Pokey and you turn yourself around. . . . In R. Gist & B. Lubin (Eds.), *Response to disaster: Psychosocial, community, and ecological approaches* (pp. 269–290). Philadelphia: Brunner/Mazel.
Gosselin, P., & Mathews, W. J. (1995). Eye movement desensitization and reprocessing in the treatment of test anxiety: A study of the effects of expectancy and eye movement. *Journal of Behavior Therapy and Experimental Psychiatry, 26*, 331–337.
Green, B., & Kimerling, R. (2004). Trauma, posttraumatic stress disorder, and health status. In I. P. Schnurr & B. Green (Eds.), *Trauma and health: Physical health consequences of exposure to extreme stress* (pp. 13–42). Washington, DC: American Psychological Association.
Griffiths, J., & Watts, R. (1992). *The Kempsey and Grafton bus crashes: The aftermath*. East Linsmore, Australia: Instructional Design Solutions.
Grunbaum, A. (1985). Explication and implications of the placebo concept. In L.

White, B. Tursky, & G. E. Schwartz (Eds.), *Placebo: Theory, research, and mechanisms* (pp. 9–36). New York: Guilford Press.

Haas, L. J., & Malouf, J. L. (1995). *Keeping up the good work: A practitioner's guide to mental health ethics*. Sarasota, FL: Professional Resource Press.

Herbert, J. D. (2000). Defining empirically supported treatments: Pitfalls and possible solutions. *Behavior Therapist, 23*, 113–134.

Herbert, J. D. (2003). The science and practice of empirically supported treatments. *Behavior Modification, 27*, 412–430.

Herbert, J. D., Lilienfeld, S. O., Lohr, J. M., Montgomery, R. W., O'Donohue, W. T., Rosen, et al. (2000). Science and pseudoscience in the development of eye movement desensitization and reprocessing: Implications for clinical psychology. *Clinical Psychology Review, 20*, 945–971.

Hines, T. (2003). *Pseudoscience and the paranormal* (2nd ed.). Amherst, NY: Prometheus Books.

Hobbs, M., Mayou, R., Harrison, B., & Worlock, P. (1996). A randomized controlled trial of psychological debriefing for victims of road traffic accidents. *British Medical Journal, 313*, 1438–1439.

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry, 69*, 621–632.

Horowitz, M. J. (1976). *Stress response syndromes*. New York: Aronson.

Huber, P. W. (1991). *Galileo's revenge: Junk science in the courtroom*. New York: Basic Books.

International Critical Incident Stress Foundation. (2012). Retrieved October 22, 2012, from *www.icisf.org*.

Ironson, G., Freund, B., Strauss, J. L., & Williams, J. (2002). Comparison of two treatments for traumatic stress: A community-based study of EMDR and prolonged exposure. *Journal of Clinical Psychology, 58*, 113–128.

Kadet, A. (2002). Good grief! *Smart Money, 11*, 108–114.

Kalal, D. M. (1999). Critical thinking in clinical practice: Pseudoscience, fad psychology and the behavior therapist. *Behavior Therapist, 22*, 81–83.

Kazdin, A. E. (2002). *Research design in clinical psychology*. New York: Harper & Row.

Kazrin, A., Durac, J., & Agteros, T. (1979). Meta-meta-analysis: A new method for evaluating therapy outcome. *Behaviour Research and Therapy, 17*, 397–399.

Keane, T. M. (1998). Psychological and behavioral treatment of posttraumatic stress disorder. In P. Nathan & J. Gorman (Eds.), *A guide to treatments that work* (pp. 398–407). Oxford, UK: Oxford University Press.

Keane, T. M., Fairbank, J. A., Caddell, J. M., & Zimering, R. T. (1989). Implosive (flooding) therapy reduces symptoms of PTSD in Vietnam combat veterans. *Behavior Therapy, 20*, 245–260.

Keane, T. M., Zimering, R. T., & Caddell, J. M. (1985). A behavioral formulation of posttraumatic stress disorder in Vietnam veterans. *Behavior Therapist*, *8*, 9–12.

Kenardy, J. A., Webster, R. A., Lewin, T. J., Carr, V. J., Hazell, P. L., & Carter, G. L. (1996). Stress debriefing and patterns of recovery following a natural disaster. *Journal of Traumatic Stress*, *9*, 37–49.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV

disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*, 593–602.
Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry, 52*, 1048–1060.
Kitchener, B. A., & Jorm, A. F. (2002). *Mental health manual*. Canberra, Australia: Center for Mental Health Research.
Klein, D. F. (2000). Flawed meta-analyses comoparing psychotherapy with pharmacotherapy. *American Journal of Psychiatry, 157*, 1204–1211.
Kring, A. M., Johnson, S., Davison, G. C., & Neale, J. (2012). *Abnormal psychology*. Hoboken, NJ: Wiley.
Lee, C. W., & Cuipers, P. (2013). A meta-analysis of the contribution of eye movements in processing emotional memories. *Journal of Behavior Therapy and Experimental Psychiatry, 44*, 21–39.
Lee, C. W., & Drummond, P. (2008). Effects of eye movement versus therapist instructions on the processing of distressing memories. *Journal of Anxiety Disorders, 22*, 801–808.
Lee, C. W., Gavriel, H., Drummond, P., Richards, J., & Greenwald, R. (2002). Treatment of PTSD: Stress inoculation training with prolonged exposure compared to EMDR. *Journal of Clinical Psychology, 58*, 1071–1089.
Lee, C. W., Slade, P., & Lygo, V. (1996). The influence of psychological debriefing on emotional adaptation in women following early miscarriage: A preliminary study. *British Journal of Medical Psychology, 69*, 47–58.
LeLorier, J., Gregoire, G., Benhaddad, A., LaPierre, J., & Derderian, F. (1997). Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine, 337*, 536–542.
Leykin, Y., & DeRubeis, R. J., (2009). Allegiance in psychotherapy outcome research: Separating association from bias. *Clinical Psychology: Science and Practice, 16*, 54–65.
Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53–70.
Lilienfeld, S. O. (2011). Distinguishing scientific from pseudoscientific psychotherapies: Evaluating the role of theoretical plausibility with a little help from Reverend Bayes. *Clinical Psychology: Science and Practice, 18*, 105–112.
Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2003). *Science and pseudoscience in clinical psychology*. New York: Guilford Press.
Litz, B. T., & Keane, T. M. (1989). Information processing in anxiety disorders: Application to the understanding of post-traumatic stress disorder. *Clinical Psychology Review, 9*, 243–257.
Liverant, G. I., Suvak, M. K., Pineles, S. L., & Resick, P. A. (2012). Changes in posttraumatic stress disorder and depressive symptoms during cognitive processing therapy: Evidence for concurrent change. *Journal of Consulting and Clinical Psychology, 80*, 957–967.
Lohr, J. M. (2001). Sakai et al. is not an adequate demonstration of TFT effectiveness. *Journal of Clinical Psychology, 57*, 1225–1235.
Lohr, J. M., DeMaio, C., & McGlynn, F. D. (2003). Specific and nonspecific treatment factors in experimental analysis of behavioral treatment efficacy. *Behavior Modification, 27*, 322–368.
Lohr, J. M., Lilienfeld, S. O., Tolin, D. F., & Herbert, J. D. (1999). Eye movement

desensitization and reprocessing (EMDR): An analysis of specific vs. nonspecific treatment factors. *Journal of Anxiety Disorders, 13*, 185–207.

Lohr, J. M., Olatunji, B. O., Parker, L., & DeMaio, C. (2005). Experimental analysis of specific treatment factors: Efficacy and practice implications. *Journal of Clinical Psychology, 61*, 819–834.

Lohr, J. M., Tolin, D. F., & Lilienfeld, S. O. (1998). Efficacy of eye movement desensitization and reprocessing: Implications for behavior therapy. *Behavior Therapy, 29*, 123–156.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95–106.

Lytle, R. A., Hazlett-Stevens, H., & Borkovec, T. D. (2002). Efficacy of eye movement desensitization in the treatment of cognitive intrusions related to a past stressful event. *Journal of Anxiety Disorders, 16*, 273–288.

Macklin, M., Metzger, L. J., Lasko, N. B., Berry, N. J., Orr, S. P., & Pitman, R. K. (2000). Five-year follow-up study of eye movement desensitization and reprocessing therapy for combat-related posttraumatic stress disorder. *Comprehensive Psychiatry, 41*, 24–27.

Mahoney, M. J. (1978). Experimental methods and outcome evaluation. *Journal of Consulting and Clinical Psychology, 46*, 660–672.

Mandler, G. (1982). Stress and thought processes. In: L. Goldberger & S. Breznitz (Eds.), *Handbook of stress: Theoretical and clinical aspects* (pp. 88–104). New York: Free Press.

Marks, I., Lovell, K., Noshirvani, H., Livanou, M., & Thrasher, S. (1998). Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring: A controlled study. *Archives of General Psychiatry, 55*, 317–325.

Masten, A. S. (1994). Resilience in individual development: Successful adaptation despite risk and adversity. In M. C. Wang & E. W. Gordon (Eds.), *Educational resilience in inner-city America: Changes and prospects* (pp. 3–25). Hillsdale, NJ: Erlbaum.

Masten, A. S., Best, K., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology, 2*, 425–444.

McCann, I. L., Sakeim, D. K., & Abramson, D. J. (1988). Trauma and victimization: A model for psychological adaptation. *Counseling Psychologist, 16*, 531–594.

McCann, J. T., Shindler, K. L., & Hammond, T. R. (2003). The science and pseudoscience of expert testimony. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 77–108). New York: Guilford Press.

McFall, R. (1991). A manifesto for the science of clinical psychology. *Clinical Psychologist, 44*, 75–88.

McNally, R. J. (1999). On eye movements and animal magnetism: A reply to Greenwald's defense of EMDR. *Journal of Anxiety Disorders, 13*, 617–620.

McNally, R. J., Bryant, R. A., & Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest, 4*, 45–79.

McNally, R. M. (2012). Are we winning the war against posttraumatic stress disorder? *Science, 336*, 872–874.

Merckelbach, H., Hogervorst, E., Kampman, M., & deJong, A. (1994). Effects of "eye movement desensitization" on emotional processing in normal subjects. *Behavioural and Cognitive Psychotherapy, 22*, 331–335.

Mitchell, J. T. (1983). When disaster strikes . . . the critical incident stress debriefing process. *Journal of Emergency Medical Services, 8*, 36–39.

Mitchell, J. T. (1988a). Development and functions of a critical incident stress debriefing team. *Journal of Emergency Medical Services, 13*, 42–46.

Mitchell, J. T. (1988b). The history, status, and future of critical incident stress debriefing. *Journal of Emergency Medical Services, 13*, 49–52.

Mitchell, J. T., & Everly, G. S., Jr. (1993). *Critical incident stress debriefing: An operations manual. The prevention of traumatic stress among emergency services and disaster workers*. Ellicott City, MD: Chevron.

Mitchell, J. T., & Everly, G. S., Jr. (1995). Critical incident stress debriefing and the prevention of work related traumatic stress among high risk occupational groups. In G. S. Everly Jr. & J. M. Lating (Eds.), *Psychotraumatology: Key papers and core concepts in post traumatic stress* (pp. 267–280). New York: Plenum Press.

Mitchell, J. T., & Everly, G. S., Jr. (1998). Critical incident stress management: A new era in crisis intervention. *Traumatic Stress Points, 12*, 6–11.

Nathan, P. E. (2000). The Boulder Model: A dream deferred or lost. *American Psychologist, 55*, 250–252.

National Fire Prevention Association. (2012). *Standard on fire department occupational safety and health program*. Accessed at *www.nfpa.org*.

National Institute for Clinical Excellence. (2004). Reviewing and grading the evidence (Chapter 7). *Guideline development methods*. London: Author.

National Institute for Clinical Excellence. (2005). *Posttraumatic stress disorder: The management of PTSD in adults and children in primary and secondary care* (Vol. 26). Trowbridge, Wiltshire, UK: Cromwell Press.

Newbold, K. M., Lohr, J. M., & Gist, R. (2008). Apprehended without warrant: Issues of evidentiary warrant for critical incident services and related trauma interventions in a federal law enforcement agency. *Criminal Justice and Behavior, 35*, 1337–1353.

Nezu, A. (1986). Efficacy of a social problem-solving therapy approach to unipolar depression. *Journal of Consulting and Clinical Psychology, 54*, 196–202.

Nezu, A., & Perri, M. G. (1989). Social problem-solving therapy for unipolar depression: An initial dismantling investigation. *Journal of Consulting and Clinical Psychology, 57*, 408–413.

Nishith, P., Nixon, R. D. V., & Resick, P. A. (2005). Resolution of trauma-related guilt following treatment of PTSD in female rape victims: A result of cognitive processing therapy targeting comorbid depression? *Journal of Affective Disorders, 86*, 259–265.

Nishith, R., Resick, P. A., & Griffin, M. G. (2002). Pattern of change in prolonged exposure and cognitive-processing therapy of female rape victims with posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 70*, 880–886.

Ohayon, M. M., & Shapiro, C. M. (2000). Sleep disturbances and psychiatric

disorders associated with posttraumatic stress disorders in the general population. *Comprehensive Psychiatry, 41*, 469–478.

Olatunji, B. O., Deacon, B. J., & Abramowitz, J. S. (2009). The cruelest cure?: Ethical issues in the implementation of exposure-based treatments. *Cognitive and Behavioral Practice, 16*, 172–180.

Oxford Center for Evidence Based Medicine. (2012). *Levels of evidence*. Retrieved May 13, 2013, from *www.cebm.net/?o=1025*.

Park, R. L. (2000). *Voodoo science: The road from foolishness to fraud*. Oxford, UK: Oxford University Press.

Parker, C. L., Everly, G. S., Jr., Barnett, D. J., & Links, J. M. (2006). Establishing evidence-informed core competencies in psychological first aid for public health personnel. *International Journal of Emergency Mental Health, 8*, 83–92.

Perkins, B. R., & Rouanzoin, C. C. (2002). A critical evaluation of current views regarding eye movement desensitization and reprocessing (EMDR): Clarifying points of confusion. *Journal of Clinical Psychology, 58*, 77–97.

Pitman, R. K., Orr, S. P., Altman, B., & Longpre, R. E. (1996). Emotional processing during eye movement desensitization and reprocessing therapy of Vietnam veterans with chronic post-traumatic stress disorder. *Comprehensive Psychiatry, 37*, 419–429.

Power, K., McGoldrick, T., Brown, K., Buchanan, R., Sharp, D., Swanson, V., et al. (2002). A controlled comparison of eye movement desensitization and reprocessing versus exposure plus cognitive restructuring versus waiting list in the treatment of post-traumatic stress disorder. *Clinical Psychology and Psychotherapy, 9*, 299–318.

Powers, M. B., Halpern, J. M., Ferenschak, M. P., Gillihan, S. J., & Foa, E. B. (2010). A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clinical Psychology Review, 30*, 635–641.

Renfrey, G., & Spates, C. (1994). Eye movement desensitization: A partial dismantling study. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 231–239.

Resick, P. A., Galovksi, Uhlmansiek, M. O., Sher, C. D., Clum, G. A., & Young-Xu, Y. (2008). A randomized clinical trial to dismantle components of cognitive processing therapy for posttraumatic stress disorder in female victims of interpersonal violence. *Journal of Consulting and Clinical Psychology, 76*, 243–258.

Resick, P. A., Nishith, P., Weaver, T. L., Astin, M. C., & Feuer, C. A. (2002). A comparison of cognitive-processing therapy and prolonged exposure and a waiting condition for the treatment of chronic posttraumatic stress disorder in female rape victims. *Journal of Consulting and Clinical Psychology, 70*, 867–879.

Resick, P. A., & Schnicke, M. K. (1992). Cognitive processing therapy for sexual assault victims. *Journal of Consulting and Clinical Psychology, 60*, 748–756.

Resick, P. A., & Schnicke, M. K. (1993). *Cognitive processing therapy for rape victims: A treatment manual*. Newbury Park, CA: Sage.

Rizvi, S. L., Vogt, D. S., & Resick, P. A. (2009). Cognitive and affective predictors of treatment outcome in cognitive processing therapy and prolonged exposure for posttraumatic stress disorder. *Behaviour Research and Therapy, 47*, 737–743.

Robinson, R. C., & Mitchell, J. T. (1993). Evaluation of psychological debriefings. *Journal of Traumatic Stress, 6*, 367–382.

Rose, S., & Bisson, J. (1999). Brief early psychological interventions following trauma: A systematic review of the literature. *Journal of Traumatic Stress, 11*, 679–710.

Rose, S., Bisson, J., & Wesselly, S. (2004). Psychological debriefing for preventing posttraumatic stress disorder. (Cochrane Review). *Cochrane Library*, *2*. (Update Software Ltd.)

Rosen, G. M. (1999). Treatment fidelity and research on eye movement desensitization and reprocessing (EMDR). *Journal of Anxiety Disorders, 13*, 173–184.

Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification, 27*, 300–312.

Rothbaum, B. O., Astin, M. C., & Marsteller, F. (2005). Prolonged exposure versus eye movement desensitization and reprocessing (EMDR) for PTSD rape victims. *Journal of Traumatic Stress, 18*, 607–616.

Russell, M. C. (2008). Scientific resistance to research, training, and utilization of eye movement desensitization and reprocessing (EMDR) therapy in treating post-war disorders. *Social Science in Medicine, 67*, 1737–1746.

Ruzek, J. I., Brymer, M. J., Jacobs, A. K., Layne, C. M., Vernberg, E. M., & Watson, P. J. (2007). Psychological first aid. *Journal of Mental Health Counseling, 29*, 17–49.

Sanchez-Meca, J., & Marin-Martinez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods, 13*, 31–48.

Sanderson, A., & Carpenter, R. (1992). Eye movement desensitization versus image confrontation: A single-session crossover study of 58 phobic subjects. *Journal of Behavior Therapy and Experimental Psychiatry, 23*, 269–275,

Schubert, S. J., Lee, C. W., & Drummond, P. (2011). The efficacy and psychophysiological correlates of dual-attention tasks in eye movement desensitization and reprocessing (EMDR). *Journal of Anxiety Disorders, 25*, 1–11.

Seidler, G. H., & Wagner, F. E. (2006). Comparing the efficacy of EMDR and trauma-focused cognitive-behavioral therapy in the treatment of PTSD: A meta-analytic study. *Psychological Medicine, 36*, 1515–1522.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14.

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology: Empirical validation of interventions. *American Psychologist, 60*, 410–421.

Shapiro, F. (1989). Efficacy of the eye movement desensitization procedure in the treatment of traumatic memories. *Journal of Traumatic Stress, 2*, 199–223.

Shapiro, F. (1995a). *Eye movement desensitization and reprocessing (EMDR): Basic principles, protocols, and procedures*. New York: Guilford Press.

Shapiro, F. (1995b, November 6). *EMDR research evaluations* [Public electronic mail publication]. Available at *traumaticstress@netcom.com*.

Shapiro, F. (1996). Eye movement desensitization and reprocessing: Evaluation of controlled PTSD research. *Journal of Behavior Therapy and Experimental Psychiatry, 27*, 209–218.

Shapiro, F. (2001). *Eye movement desensitization and reprocessing (EMDR):*

Basic principles, protocols, and procedures (2nd ed.). New York: Guilford Press.

Shapiro, F. (2002). EMDR 12 years after its introduction: Past and future research. *Journal of Clinical Psychology, 58*, 1–22.

Shapiro, R. (2005). *EMDR solutions: Pathways to healing*. New York: Norton.

Sharpley, J. G., Fear, N. T., Greenberg, M., Jones, M., & Wessely, S. (2008). Predeployment stress briefing: Does it have an effect? *Occupational Medicine, 58*, 30–34.

Sinclair, C., Simon, N. P., & Pettifor, J. L. (1996). The history of ethical codes and licensure. In L. J. Bass, S. T. DeMers, J. R. P. Ogloff, C. Peterson, J. L. Pettifor, R. P. Reaves, et al. (Eds.), *Professional conduct and discipline in psychology* (pp. 1–15). Washington, DC: American Psychological Association.

Singer, M., & Lalich, J. (1996). *Crazy therapies: What are they? Do they work?* San Francisco: Jossey-Bass.

Solomon, R. M., & Shapiro, F. (2008). EMDR and the adaptive information processing model: Potential mechanisms of change. *Journal of EMDR Practice and Research, 2*, 315–325.

Spates, C. R., Koch, E., Cusack, K., Pagoto, S., & Waller, S. (2009). Eye movement desensitization and reprocessing. In E. B. Foa, T. M. Keane, M. J. Friedman, & J. A. Cohen (Eds.), *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies*. New York: Guilford Press.

Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 433–455). New York: Oxford University Press.

Steenkamp, M. M., Nash, W. P., & Litz, B. T. (2013). Post-traumatic stress disorder: Review of the comprehensive soldier fitness program. *American Journal of Preventive Medicine, 44*, 507–512.

Stuart, R. B., & Lilienfeld, S. O. (2007). The evidence missing from evidence-based practice. *American Psychologist, 62*, 615–616.

Stuhlmiller, C., & Dunning, C. (2000). Challenging the mainstream: From pathogenic to salutogenic models of posttrauma intervention. In J. M. Violanti, D. Paton, & C. Dunning (Eds.), *Posttraumatic stress intervention: Challenges, issues, and perspectives* (pp. 10–42). Springfield, IL: Charles C Thomas.

Tallis, F., & Smith, E. (1994). Does rapid eye movement desensitization facilitate emotional processing? *Behaviour Research and Therapy*, *32*, 459–461.

Tarrier, N., Pilgrim, H., Sommerfield, C., Faragher, B., Reynolds, M., Graham, E., et al. (1999a). A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 67*, 13–18.

Tarrier, N., & Sommerfield, C. (2004). Treatment of chronic PTSD by cognitive therapy and imaginal exposure: 5-year follow-up. *Behavior Therapy, 35*, 231–246.

Tarrier, N., Sommerfield, C., Pilgrim, H., & Humphreys, L. (1999b). Cognitive therapy or imaginal exposure in the treatment of posttraumatic stress disorder: Twelve-month follow-up. *British Journal of Psychiatry, 175*, 571–575.

Task Force on Promotion and Dissemination Psychological Procedures. (1995). Training and dissemination of empirically validated treatments. *Clinical Psychologist, 48*, 3–23.

Taylor, S., Thordarson, D. S., Maxfield, L., Fedoroff, I. C., Lovell, K., & Ogrodniczuk, J. (2003). Comparative efficacy, speed, and adverse effects of three PTSD treatments: Exposure therapy, EMDR, and relaxation training. *Journal of Consulting and Clinical Psychology, 71*, 330–338.

Tedeschi, R. G., & McNally, R. J. (2011). Can we facilitate posttraumatic growth in combat veterans. *American Psychologist, 66*, 19–24.

Tolin, D. F. (2010). Is cognitive-behavioral therapy more effective than other therapies?: A meta-analytic review. *Clinical Psychology Review, 30,* 710–720.

Tryon, W. W. (2005). Possible mechanisms for why desensitization and exposure therapy work. *Clinical Psychology Review, 25*, 67–95.

van den Hout, M. A., & Englehard, I. M. (2012). How does EMDR work? *Journal of Experimental Psychopathology, 3*, 724–728.

van Emmerik, A. A. P., Kamphuis, J. H., Hulsbosch, A. M., & Emmelkamp, P. M. G. (2002). Single session debriefing after psychological trauma: A meta-analysis. *The Lancet, 360*, 776–771.

Varker, T., & Devilly, G. J. (2012). An analogue trail of inoculation/resilience training for emergency services personnel: Proof of concept. *Journal of Anxiety Disorders, 26*, 696–701.

Vernberg, E. M., Steinberg, A. M., Jacobs, A. K., Brymer, M. J., Watson, P, Osofsky, J. D., et al. (2008). Innovations in disaster mental health. *Professional Psychology: Research and Practice, 39*, 381–388.

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings.* Mahwah, NJ: Erlbaum.

Wilson, D. L., Silver, S. M., Covi, W. G., & Foster, S. (1996). Eye movement desensitization and reprocessing: Effectiveness and autonomic correlates. *Journal of Behavior Therapy and Experimental Psychiatry, 27*, 219–229.

Wilson, N. (2003). Commercializing mental health issues: Entertainment, advertising, and psychological advice. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 425–460). New York: Guilford Press.

World Health Organization. (2006). *Single session psychological debriefing: Not recommended*. Retrieved August 1, 2007, from *www.who/hac/techguidance/pht/13643.pdf.*

World Health Organization. (2011). *Psychological first aid: Guide for field workers*. Geneva, Switzerland: Author.

CHAPTER ELEVEN

Controversial Treatments for Alcohol Use Disorders

James MacKillop and Joshua C. Gray

INTRODUCTION

The modern array of treatment options and facilities for alcohol use disorders (AUDs) belies the fact that formal treatment for alcohol problems is a relatively recent endeavor. Historically, excessive drinking was typically viewed through a moral lens as a character defect, reflecting weakened mental capacity and a lack of self-control (Levine, 1978). As a result, societal responses were primarily in the form of legal sanctions, such as fines or jail sentences, and intercession by religious leaders for spiritual and moral restoration. During the 19th and early 20th centuries, grass-roots movements, such as the Washingtonians and, later, Alcoholics Anonymous, also arose to aid individuals with drinking problems (for a full review, see White, 1998).

Over the course of the 20th century, treatment for alcohol problems shifted into the purview of medicine, clinical psychology, and other clinical disciplines for a number of reasons. Psychology, physiology, pharmacology, and neuroscience burgeoned as scientific disciplines and offered new theories for the causes of alcoholism (see Miller & Hester, 1995, for a review of 13 theoretical perspectives) and novel ways to parse the condition (e.g., Jellinek, 1946, 1960). Experimental methods were developed that permitted the systematic study of alcohol consumption under controlled conditions (Lisman, 1974; Mello & Mendelson, 1965), further strengthening the scientific enterprise. Perhaps the single most important event in advancing a scientific approach to AUD treatment in the United States, however, was passage of the Comprehensive Alcohol Abuse and Alcoholism Prevention,

Treatment, and Rehabilitation Act in 1970. This act in turn created the National Institute of Alcohol Abuse and Alcoholism (NIAAA) in 1971, which was in large part responsible for the substantial expansion of AUD clinical research. Furthermore, it marked the inception of the professional treatment establishment from which inpatient and outpatient clinics subsequently proliferated.

The professional treatment establishment that now exists, however, is highly heterogeneous and fractious in terms of perspectives on best-practices. This friction in part stems from the fact that the alcohol treatment enterprise includes two broad constituencies: treatment providers who have limited formal training but have personally experienced AUDs (or other addictive disorders) and are now stably recovered; and treatment providers who are formally trained but have not personally suffered from these conditions. For example, in two recent studies of treatment professionals, approximately half of the treatment providers were successfully recovered individuals (Curtis & Eby, 2010; Davis & Rosenberg, 2012). A heuristic distinction that has been made between these groups is that of "craftsmen," the informally trained recovered individuals, and "scientists," the formally trained clinicians (Kalb & Propper, 1976). The craftsmen are often characterized (or perhaps caricatured) as having strong and uncritical adherence and advocacy of specific approaches, with relatively little interest in systematic research, whereas the scientists are characterized as being overly dogmatic and unwilling to consider the merits of treatments that are not directly supported by empirical evidence. Although recent data suggest that this distinction is increasingly less black-and-white (Davis & Rosenberg, 2012), the fact remains that professional alcohol treatment continues to comprise two constituencies with quite different perspectives. Based on the varying emphasis on research, a wide assortment of treatment approaches are currently employed, with considerable controversy and highly variable levels of empirical support.

Reviewing the controversial treatments for AUDs is the first goal of this chapter. We define treatments as "controversial" if they have become widely adopted or have generated popular appeal, but have been demonstrated by controlled research to be of questionable or even nonexistent efficacy. Equally, other controversial treatments are defined by the inverse: approaches that have been shown to be effective but remain largely unused due to practitioner taboos. In other words, controversial treatments are those that are in use despite the lack of evidence for their efficacy or are not in use despite evidence for their efficacy. Specifically, the controversial approaches we discuss are confrontational Johnson Institute "interventions;" Alcoholics Anonymous, and other mutual-help groups; the prevention program Drug Abuse Resistance Education; and interventions designed to promote moderate drinking. The second goal of the chapter is to review evidence-based treatments for AUDs. Increasing the use of treatments that have foundations in behavioral science and are supported by

scientific evidence requires persistently disseminating information about these interventions and the supporting evidence. Here, we will review reinforcement-based approaches, cognitive-behavioral approaches, motivational interviewing, marital and family therapy, brief interventions, and pharmacological treatment. In a concluding section, we reflect on the state of the field and future directions for a scientifically based approach to treating AUDs.

CONTROVERSIAL TREATMENTS

The Johnson Intervention

If asked, most members of the general public would almost certainly report "an intervention" as the most effective strategy for motivating an individual to seek treatment for an AUD. This refers to a structured confrontation of the individual by family and friends to convince the person to stop drinking and seek treatment immediately. It was initially developed in the 1970s by Vernon Johnson, an Episcopal priest and founder of an alcohol treatment facility, the Johnson Institute, giving rise to the strategy's formal name, the Johnson Intervention. The format of the Johnson Intervention is relatively well known. Responding to a false pretext, targeted individuals are lured to a meeting with the important people in their lives. There, they are systematically confronted with the negative effects of their drinking (or drug use) and the consequences that will follow if they elect not to immediately accept the offer of treatment. The putative strategy is to overwhelm the individuals' denial, thereby motivating them for treatment, and then initiate the process (Johnson, 1986; Twerski, 1983).

The Johnson Intervention is not the only form of confrontational intervention used in AUD treatment (White & Miller, 2007), but it has become the most widespread for a number of reasons. A Johnson Intervention was the catalyst for former U.S. First Lady Betty Ford (wife of President Gerald Ford) to seek treatment, bringing it into the popular consciousness. Moreover, because Johnson Interventions have the potential for high drama, they have increasingly become plot lines on a wide variety of popular television programs (e.g., *The Sopranos*, *Party of Five*, *Beverly Hills 90210*). Indeed, since 2005, an eponymous reality television program, *Intervention* (A&E, 2012), has followed over 200 individuals over the course of a Johnson Intervention, winning an Emmy award in 2009.

The proponents of Johnson Interventions make undocumented claims of very high rates of successful long-term recovery (e.g., 80%; Assistance in Recovery, 2012). From the perspective of controlled clinical research, however, the Johnson Intervention is not robustly supported. Liepman, Nirenberg, and Begin (1989) examined the effects of Johnson Intervention on treatment initiation and reported high rates of entry, but acknowledged several limitations, including very high rates of families who did

not undertake the Johnson Intervention, nonrandom assignment of participants, and a small sample size. In another study, Loneck, Garrett, and Banks (1996a) found that, compared with four other strategies, individuals who underwent a Johnson Intervention were more likely to enter treatment. However, their methodology was retrospective and only represented the cases in which the family members followed through with the Johnson Intervention. Furthermore, participants in the Johnson Intervention condition were more likely to relapse than those in three of the four comparison groups (Loneck, Garrett, & Banks, 1996b). More recently, Miller, Meyers, and Tonigan (1999) conducted a more methodologically rigorous study and found that family compliance rates for participation in an intervention were strikingly low (30%) compared with another intervention condition and an Alcoholics Anonymous condition. In addition to overstated claims of efficacy, Johnson Interventions may in fact cause harm, although research substantiating this possibility is lacking. Implicit in the preceding findings is that if the Johnson Intervention is believed to be the only option for motivating a person to seek treatment, most families are unwilling to undertake the ordeal. For those that are, coercive techniques that confront individuals with ultimata and cataclysmic consequences, such as the dissolution of a marriage, may elicit extremely negative reactions (Galanter, 1993). Furthermore, confrontational Johnson Interventions do not consider the possibility of exacerbating commonly comorbid conditions, such as major depression or posttraumatic stress disorder.

The last negative consequence of Johnson Interventions comes in the form of opportunity cost. Johnson Interventions may displace the implementation of alternative strategies for which a strong evidence base exists. The best supported is the Community Reinforcement Approach Family Training (CRAFT; Meyers, Miller, Hill, & Tonigan, 1998; Meyers & Smith, 1997), which has foundations in operant theory and is discussed in detail later in this chapter. The second is a more gradual three-step approach, A Relational Sequence for Engagement (ARISE; Garrett et al., 1998), which starts with social and emotional support from family and friends, and progresses to a typical Johnson Intervention while striving to minimize adverse reactions from the targeted individual. During Stage 1, a family member or friend calls all involved individuals to set up the intervention network and a meeting. The alcohol user is informed and encouraged to attend the meeting. In Stage 2, the family meets regardless of the attendance of the alcohol user, and they assess the alcohol problem in the context of the family system. If the alcohol user continues to avoid these meetings, then Stage 3 is initiated, whereby the traditional confrontational Johnson Intervention approach is employed. In a within-subjects clinical trial, Landau et al. (2004) found that the ARISE approach resulted in 83% of the participants engaging in individual treatment or self-help. Critically, however, 80% were successful in the nonconfrontational Stages 1 and 2 and only 3% at Stage 3. Although this is the only controlled study on ARISE for AUDs, it

is a promising start and suggests the importance of actively engaging the individual in discussing treatment.

Taken together, the Johnson Intervention is an exemplar of a pseudoscientific treatment for AUDs. It is in common practice, vocally advocated by adherents who cite very positive outcomes from unverifiable and nonscientific sources, and not supported in controlled research. Furthermore, as it may potentially result in adverse consequences for its targets and at least one evidence-based practice is available as an alternative. In sum, although the melodrama of surprising and confronting individuals with AUDs and other addictive disorders appears to make for good television, the evidence suggests that it does not make for good treatment.

ALCOHOLICS ANONYMOUS

Founded in 1935, with a current estimated 114,000 groups worldwide and over 2,000,000 members, Alcoholics Anonymous (AA) is almost certainly the largest mutual-help group in the world (Alcoholics Anonymous, 2012). Despite its size, AA is flat in structure. The governing body is only one level above the individual chapters and cannot make decisions that affect AA as a whole. For this reason, AA remains both ideologically and organizationally very close to its grass-roots origins. Like the Johnson Intervention, the general format of AA is also familiar to the general public. Individuals meet together for fellowship, share their past experiences with alcohol with a nonjudgmental group of fellow sufferers, and try to follow the 12 steps toward individual and interpersonal rehabilitation. A religious component, though nondenominational, is central; seven of the steps refer specifically to God, a "Higher Power," or Him. A sponsorship system is often in place, in which a new member is mentored by a stable senior member. Additionally, newcomers are advised to immerse themselves by attending a meeting each night for the first 90 days.

AA is perhaps the most controversial AUD treatment (Kaskutas, 2009). Members and proponents of AA are often perceived as overly theistic and optimistic, and as unaccepting of other perspectives (Ellis & Schoenfeld, 1990; Tournier, 1979); some have gone as far as to compare AA to a cult (Bufe, 1991). The AA etiological model, not having been updated in over 50 years, has also been criticized as highly outdated (Kelly, 2013). The treatment model has been criticized for fostering dependence on external factors that are beyond the individual's control (Ellis & Schoenfeld, 1990) and for applying a one-size-fits-all approach (Kelly, 2013). Although AA is not affiliated with any professional organization or treatment, many formal treatment programs have nonetheless adopted its perspective (Roman & Johnson, 1998). Conversely, research-trained clinicians have been criticized for being myopic and discounting the potential of AA and other self-help treatments (Chiauzzi & Liljegren, 1993).

The gold standard for empirical support for the efficacy of clinical interventions is the randomized controlled trial (RCT), but relatively few studies have examined AA in RCTs. Of those that have, the data are not especially supportive. For example, three randomized trials of AA have been undertaken involving populations mandated to attend AA and were not supportive (Alford, 1980; Ditman, 1967; Walsh et al., 1991). Meta-analyses and systematic reviews of controlled studies have also concluded that the balance of evidence does not support the efficacy of AA (Ferri, Amato, & Davoli, 2006; Kownacki & Shadish, 1999).

Superficially, these findings suggest that AA is indeed a pseudoscientific treatment, persisting by virtue of intuitive appeal and strident adherents despite weak empirical support. However, there are a number of reasons that this conclusion may not be accurate, and there remains a need to consider additional research evidence. First, as a consumer-led nonprofessional mutual-help organization, a number of AA's characteristics distinguish it from formal psychological or medical interventions. To start, AA does not purport to be a formal treatment per se (Kelly, 2013) and, although the AA format is well known, the nature of AA groups and the experiences an individual has in AA may be highly variable. All that is required for an AA group is that two or more alcoholics meet together for the purpose of achieving sobriety (McCrady, Horvath, & Delaney, 2003). As such, none of the essential features of formal treatment outcome research, like treatment manuals and fidelity checks, can be applied. Random assignment of individuals to AA does not replicate the typical pathway that leads to participation and has the potential to create a negatively predisposed group of subjects (McCrady et al., 2003). Furthermore, random assignment to conditions other than AA may be ethically questionable, as it involves prohibiting individuals from participating in a free community resource (Kelly, 2013).

By virtue of these innate characteristics, quasi-experimental and naturalistic studies may be more appropriate for understanding the clinical utility of AA. Here, the data are persuasive that AA participation is associated with improved drinking outcomes. For example, in a large naturalistic study on individuals in the Veterans Administration, Ouimette, Moos, and Finney (1998) found that, relative to individuals who only received outpatient treatment, patients who attended only AA meetings were more likely to be abstinent at 1-year follow-up, although, not surprisingly, the best outcomes were for individuals attending both treatment modalities. Timko, Moos, Finney, and Lesar (2000) studied the outcome of 248 problem drinkers who self-selected AA, formal treatment, formal treatment plus AA, or no treatment at all at 1-year, 3-year, and 8-year follow-up. Those who chose no treatment fared the worst, and the AA-only and combined AA/formal treatment groups fared better than the formal treatment group alone at 1 and 3 years, although not at the 8-year follow-up. A follow-up at 16 years found that the duration of AA attendance in years 1 to 3

independently predicted abstinence and lower drinking problems (Moos & Moos, 2006).

Another source of data that bears on AA comes from studies that do not focus on AA per se, but on facilitating participation in a 12-step program. This approach, referred to as 12-step facilitation (TSF), is formalized in a way that is compatible with other psychological interventions and is more amenable to controlled study. It is conceptually grounded in the tenets of AA and other 12-step groups and attempts to foster a commitment to participate in AA (Project MATCH Research Group, 1997). In the large multisite clinical trial, Matching Alcohol Treatments to Client Heterogeneity (Project MATCH), TSF was contrasted with cognitive-behavioral coping skills management and motivational enhancement therapy. Although TSF intervention was not a test of AA as a treatment, the three treatments produced comparable outcomes (Project MATCH Research Group, 1998). The lack of significant outcome differences and retention rates among the three treatment conditions suggests that TSF is equivalent to evidence-based treatments. In a study comparing relapse prevention and TSF following inpatient treatment, both interventions had equally positive impacts, although women, polysubstance users, and individuals who were high in psychological distress benefited more from TSF (Brown, Seraganian, Tremblay, & Annis, 2002). These findings converge with previous evidence that individuals with AUDs who are able to choose treatment pathways have better outcomes (Sanchez-Craig & Lei, 1986). Taken together, it appears that among people who self-select into AA, there are significant short- and long-term benefits.

The empirical literature on AA has also become considerably more sophisticated in identifying the factors that predict whether a person is well suited for AA and the mechanisms that underlie AA's positive effects. For example, in Project MATCH, patient ratings of AA's conceptual and spiritual underpinnings predicted meeting attendance, engagement in AA practices, and greater abstinence (Tonigan, Miller, & Connors, 2000). Similarly, although nonreligious individuals attend AA (Tonigan, Miller, & Schermer, 2002), they attend less frequently (Tonigan et al., 2002) and are more likely to drop out (Kelly & Moos, 2003). With regard to mechanisms, a recent review of 19 studies found that several potential mechanisms, including increased self-efficacy, increased coping skills, and the development of a more adaptive social network, appeared to be responsible for AA's positive effects (Kelly, Magill, & Stout, 2009). These social network changes include both decreasing contact with pro-drinking individuals and increasing contact with pro-abstinence individuals. In contrast, there was limited evidence for the spiritual mechanisms or AA-specific practices being the mechanisms of action. This finding converges with a previous review suggesting that social network changes are substantively responsible for positive AA influences (Groh, Jason, & Keys, 2008). In a recent study examining mechanisms of AA attendance effects in Project MATCH

(Kelly, Hoeppner, Stout, & Pagano, 2012), there was consistent evidence that adaptive social network changes partially mediated the link between AA attendance and positive drinking outcomes.

These selection and mechanism findings are important for both obvious and more subtle reasons. Clearly, knowing client characteristics associated with positive AA outcomes can inform clinicians with regard to which patients may be most appropriate for AA. Similarly, understanding AA's mechanisms is important for clinicians to anticipate probable positive consequences from AA and to orient recommendations toward those processes. Furthermore, understanding AA's mechanisms of behavior change, and especially the fact that these mechanisms do not appear to be content specific, may also inform more formal treatments, such as the development of interventions to foster more adaptive social networks. Although the largest by far, it is notable that AA is not the only mutual-help group for AUDs. Others include Secular Organizations for Sobriety, which shares similar tenets to AA but has no spiritual dimension; SMART recovery, which is a support group that includes elements of motivational interviewing, cognitive-behavioral therapy, and rational emotive behavior therapy; and Moderation Management, which focuses on helping people achieve nonabstinent outcomes. These mechanistic findings suggest that the individuals with AUDs may benefit from diverse types of mutual-help groups that commonly enhance self-efficacy, increase coping skills, and nurture the development of a nondrinking social network.

A final consideration with regard to AA in AUD treatment is logistical. In a mental health treatment environment that is highly focused on the costs of services, AA is unarguably a highly cost-effective resource, for it is free and self-supporting. Moreover, there is increasing evidence that participation in AA reduces general health care costs (Humphreys & Moos, 2001, 2007; Mundt, Parthasarathy, Chi, Sterling, & Campbell, 2012). For example, in a recent longitudinal study of 403 adolescents with AUDs, mutual-help group attendance following formal treatment was associated with a 4.7% health care cost reduction during a 7-year follow-up (Mundt et al., 2012). That translated into a $145 reduction in health care costs for each meeting attended, primarily from additional treatment costs, hospitalizations, and psychiatric follow-up. In addition to being free, a further advantage is that AA and other mutual-help groups are highly accessible to the public (e.g., no formal initiation, daily meetings).

To conclude, the role of AA in the science-based AUD treatment enterprise has been highly controversial, but it has been increasingly informed by data. There is weak evidence for the clinical efficacy of AA in terms of RCTs, but the RCT may mischaracterize the nature of AA and oversimplify the question at hand. There is consistent evidence that AA involvement and AA facilitation are associated with positive AUD outcomes. Moreover, research clarifying the active ingredients of AA offers novel insights that suggest common process factors are operative and may be applicable to

formal treatment. To be sure, from a scientific standpoint, there are reasons to be critical of AA's outmoded etiological model and to question the strong identification of formal treatment programs with AA principles (Kelly, 2013). Participation in community mutual-help groups like AA will not be for all patients, but, for some, AA may very well enhance formal treatment efforts. Therefore, we propose that the most appropriate approach for clinicians is to leverage AA's wide proliferation and availability, but to do so in ways that are informed by the expanding empirical literature in the area. In other words, keep the baby, throw out the bathwater.

Controlled Drinking as a Treatment Outcome

In contrast to the preceding controversies, the posttreatment goal of controlled drinking (i.e., occasional nonharmful alcohol use) is not scientifically controversial because of limited or contradictory evidence of its efficacy. To the contrary, controlled drinking has a long history of controversy, despite having an evidence base *supporting* its efficacy. The pseudoscience here is its reflexive rejection as a treatment goal and its very limited integration into most treatment programs.

Controlled drinking seeks to assist individuals with drinking problems to establish or reestablish control or moderation in drinking, typically defined as a limit on the amount and frequency of consumption, as well as avoidance of legal, social, and physical problems. This goal may not seem controversial, but it is antithetical to the "one-drink, one-drunk" principles of AA and, as most AUD treatment facilities identify with that orientation (Roman & Johnson, 1998), it also runs counter to most treatment regimens. For this reason, when clinical research first supported the prospect of controlled drinking in the early 1970s, a fractious public debate followed.

The initial data that sparked an interest in controlled drinking came from a long-term follow-up of a cohort of alcohol-dependent individuals that reported approximately 8% had resumed drinking but at healthy levels (Davies, 1962). This was not the only report of this finding, but it was important because the data were framed as the basis for questioning the assumption of permanent abstinence as the only viable treatment outcome. Subsequently, the question of controlled drinking as an acceptable treatment option erupted in controversy as a result of two studies. The first was an investigation applying behavior therapy to alcohol dependence, with the aim of achieving controlled drinking (Sobell & Sobell, 1973, 1976). The second was a longitudinal assessment of almost 600 alcohol-dependent individuals who had been treated at federally funded treatment centers (Armor, Polich, & Stambul, 1976), finding the presence of "normal" drinkers in the follow-up samples. Specifically, 12% and 22% of individuals were considered normal drinkers at 6- and 18- month follow-ups, respectively.

These data were widely publicized and attacked on scientific, political, and ethical grounds by those advocating abstinence as the only viable

treatment goal (for reviews of the controversy, see Marlatt, Larimer, Baer, & Quigley, 1993; Roizen, 1987; Sobell & Sobell, 1995). This uproar included sensational media coverage and, in the case of the Sobels, efforts to discredit the investigators and formal investigations into scientific wrongdoing. Ultimately, it was determined that no scientific misconduct had taken place—the data were what they were—but already significant damage was done. The virulent opposition and public outcry had a chilling effect on research on controlled drinking. The media that had highlighted the initial reactions to the studies and claims of misconduct largely ignored the findings of no wrongdoing (Marlatt et al., 1993). Even into the 1990s, a survey of almost 200 treatment programs found that 75% viewed controlled drinking as an unacceptable treatment goal (Rosenberg & Davis, 1994).

Nonetheless, studies reporting controlled drinking have continued to emerge, from treatment outcome studies (Miller, Walters, & Bennett, 2001), reports of self-treatment or natural recovery (Sobell, Cunningham, & Sobell, 1996), and longitudinal studies (Vaillant, 1996). A particularly impressive example is a follow-up of over 8,000 alcohol-dependent patients enrolled in large multisite clinical trials (Miller et al., 2001). This study revealed that approximately 25% of patients successfully achieved long-term abstinence and about 10% achieved long-term successful moderation. For the remaining two-thirds, the common profile was continued drinking but with substantial reductions in the amount of alcohol consumed (87%) and alcohol-related problems (60% reduction). The prospect of controlled drinking is also filtering increasingly into the popular media. The volume *Sober for Good* (Fletcher, 2002) details the narratives of over 200 individuals who successfully overcame alcohol dependence, approximately 10% of whom reported controlled drinking. Thus, these data have become harder for the scientific community, the treatment provider community, and the public to ignore.

Beyond evidence that some individuals ultimately achieve controlled drinking outcomes, the clinical research on methods for achieving this goal has been steadily growing (Saladin & Santa Ana, 2004) and an array of science-based manuals and programs are available (e.g., Miller & Muñoz, 2004; Rotgers, Kern, & Hoeltzel, 2002). For example, in a meta-analysis of 17 randomized controlled trials investigating behavioral self-control training (BSCT), a controlled drinking treatment, BSCT was found to demonstrate larger decreases in alcohol consumption and problems compared with alternative moderation-oriented interventions or no intervention, and to be equal in efficacy to abstinence-based treatments (Walters, 2001). A recent study revealed that heavy drinkers who received a computer-based moderation intervention experienced significant reduction in drinking behavior comparable with in-person moderation treatment (Hester & Delaney, 1997; Hester, Delaney, & Campbell, 2011). An additional consideration of a moderation-based approach is its attractiveness to individuals who may otherwise not seek treatment. One study of 742 treatment-seeking

drinkers found that 54% expressed preference for abstinence and 46% expressed preference for nonabstinence (Heather, Adamson, Raistrick, & Slegg, 2010). Furthermore, there is evidence that better outcomes are achieved, both for moderation and for abstinence, when problem drinkers are allowed to select their treatment pathway (e.g., Orford & Keddie, 1986; Sanchez-Craig & Lei, 1986).

Despite the robust evidence that some individuals achieve controlled drinking and the increasing evolution in perceptions of controlled drinking, it is important not to inadvertently push the pendulum in the other direction too far. In the data reviewed here, although some individuals achieved controlled drinking, it was often a relatively small proportion (e.g., Miller et al., 2001). It is clearly not a realistic goal for all individuals seeking AUD treatment. Indeed, a critical question is what characteristics predict success with controlled drinking. Rosenberg (1993) surveyed the literature and identified a number of variables that predict successful controlled drinking, including lower severity of dependence, a belief that moderation is possible, employment, younger age, psychological and social stability, and female gender. Additionally, Miller, Leckman, Delaney, and Tinkcom (1992) found that problem severity, rejection of an abstinence goal, and refusal to label oneself as an alcoholic predicted successful controlled drinking. More recently, Heather and Dawe (2005) found that severity was not predictive of success in a moderation-oriented intervention, but that impaired control over alcohol was associated with a negative outcome.

Taken together, the issue of controlled drinking in AUD treatment represents another clear example of science versus pseudoscience. It is a matter of data versus dogma. The data suggest that, as an outcome, controlled drinking happens and that controlled drinking interventions are efficacious in helping patients achieve that goal. As such, these interventions should be part of the AUD clinical enterprise, probably as part of a stepped-care framework in which it is available to individuals who have lower severity and are seeking a nonabstinence outcome (Sobell & Sobell, 1995). Perhaps most promisingly, in a recent study of over 900 members of the National Association of Alcoholism and Drug Addictions Counselors, 50% of respondents viewed controlled drinking as an acceptable treatment goal, which was a substantial increase from 25% in 1994 (Davis & Rosenberg, 2012). For many clinicians, however, the refrain that abstinence is the only viable outcome remains and, unfortunately, data do not matter to dogma.

Drug Abuse Resistance Education

Although the primary focus of this chapter is on treatment, we also review the efficacy of a prevention program, Drug Abuse Resistance Education

(DARE), because it is both controversial and an exemplar of a largely pseudoscientific intervention. Developed by the Los Angeles Police Department in 1983, DARE uses uniformed police officers to teach an educational drug abuse prevention program that focuses on negative aspects of substance abuse and emphasizes the positive aspects of a healthy lifestyle (Koch, 1994). The core curriculum is delivered to 5th- and 6th-grade students in 1-hour sessions for 17 consecutive weeks. Other components have been developed for students beginning from kindergarten through high school, creating a comprehensive developmental prevention program (Koch, 1994). Having police officers in the classroom to warn children of the dangers of using alcohol and other drugs clearly has an intuitive appeal (Clayton, Leukefeld, Harrington, & Cattarello, 1996); the program is also popular with both parents and administrators (Donnermeyer, 2000; Donnermeyer & Wurschmidt, 1997). For many schools across the United States, it became the prevention program of choice for these reasons.

Despite its best intentions, appeal, and widespread adoption, the research literature reveals an absence of evidence that DARE has any appreciable effect on preventing substance use. Large studies with ample statistical power have revealed no short-term benefits (Dukes, Ullman, & Stein, 1996; Rosenbaum, Flewelling, Bailey, Ringwalt, & Wilkinson, 1994). In addition, a meta-analysis of eight outcome studies indicated negligible effect sizes of DARE on alcohol and drug use (Ennett, Tobler, Ringwalt, & Flewelling, 1994). A more recent meta-analysis similarly revealed negligible effects (Pan & Bai, 2009). In terms of extended outcomes, studies with 6-year or longer follow-ups have found no beneficial effects on substance use (Lynam et al., 1999; Rosenbaum & Hanson, 1998). Indeed, in one study, DARE was associated with significant *increases* in substance use (Rosenbaum & Hanson, 1998). Based on these converging findings, DARE has been identified as an ineffective prevention program by the U.S. Surgeon General (Satcher, 2001) and as a potentially harmful treatment (Lilienfeld, 2007).

Importantly, like the Johnson Intervention, it appears that the negative consequences of DARE are twofold: first, the delivery of an ineffective and essentially symbolic prevention program in communities; and, second, the opportunity cost of allocating resources that could be used for evidence-based prevention programs (Coombs & Ziedonis, 1995). Despite this evidence of negligible or potentially iatrogenic effects, perhaps the greatest surprise is that the DARE program persists and continues to be implemented. This may well be because it has an extensive infrastructure and support system for interested communities (Merrill, Pinsky, Killeya-Jones, Sloboda, & Dilascio, 2006), creating powerful momentum and relative ease for new programs to be initiated. For example, despite the preceding evidence, in its most recent annual report, DARE reported the initiation of 300 new programs from 2007 to 2009 (DARE, 2009).

EVIDENCE-BASED TREATMENTS: THE STATE OF THE SCIENCE

The dissemination of research findings and evidence-based practice guidelines is critical for eliminating the disconnect between typical clinical practices and best-practices. Therefore, in the second portion of the chapter we review the primary forms of evidence-based treatment for AUDs, both psychological and pharmacological. In each case, we review their theoretical foundation, typical form of implementation, and empirical basis.

Reinforcement-Based Treatment for AUDs

An extensive scientific literature documents the application of operant learning theory to AUDs and other addictive disorders (for reviews, see Bigelow, 2001; Higgins, Heil, & Lussier, 2004). This approach has its foundations in the systematic study of the interactions between organisms and their environments, with specific emphasis on quantifying the contingencies of reinforcement and punishment in place (Skinner, 1938/1991). Alcohol and other addictive drugs are particularly amenable to this perspective because they are potent and prototypic positive and negative reinforcers. In other words, they are sought for the high they provide (i.e., positive reinforcement) and for the distress they alleviate (i.e., negative reinforcement). As such, AUDs can be conceptualized as a combination of an excessive sensitivity to the positively and negatively reinforcing effects of alcohol and a relative insensitivity to the punishing consequences of persistent drinking. A reinforcement-based perspective on AUDs has given rise to a number of clinical interventions of highly varying scope.

The first reinforcement-based treatment from this perspective was the Community Reinforcement Approach (CRA; Hunt & Azrin, 1973), which attempts to restructure the environmental contingencies in the patient's life so that abstinence becomes more rewarding than drinking. A common feature of AUDs is the vicious cycle in which excessive drinking leads to negative outcomes, such as the loss of a job or personal relationships, which in turn results in fewer reinforcing opportunities other than alcohol for the individual. In this way, people with AUDs putatively "paint themselves into a corner" from a reinforcement standpoint, with drinking being both the source of their problems and the only remaining source of solace.

To address this issue, the CRA works with the individual to restore healthy alternative forms of reinforcement that are mutually exclusive with drinking. The goal is to develop alternative reinforcers to successfully "compete" with alcohol and reduce the reinforcing value of drinking. This approach entails a macroscale perspective on the individual's life and a multifaceted rearrangement of the individual's vocational, social, recreational, and familial contingencies (for a review, see Meyers & Smith, 1995). The CRA approach begins with an idiographic assessment of the environmental contingencies that shape an individual's behavior. Use of a structured

protocol facilitates identification of external factors, such as where, when, and with whom an individual drinks; internal triggers, such as cognitions, emotions, or physiological experiences that precede drinking; and antecedents and consequents of both drinking and nondrinking behaviors (Smith & Meyers, 1995). Based on this assessment, different modules from a menu of treatment options are selected, all of which are intended to increase the amount of positive reinforcement from nondrinking sources. For example, one module is job skills counseling, which focuses on improving skills to secure and maintain work, and another is social recreational counseling, which focuses on developing nondrinking social relationships.

There is extensive scientific support for the CRA. The earliest studies reported very positive outcomes compared with standard care (Azrin, 1976; Hunt & Azrin, 1973), for both reductions in drinking and engagement in salutary alternative behaviors. For example, in one early study, individuals in the CRA condition drank an average of only 2% of posttreatment days, compared with 55% of days for individuals in the control group, and these gains were sustained at a 2-year follow-up, when 90% of the CRA group was still abstinent (Azrin, 1976). Positive outcomes have been found in a diversity of populations, including challenging groups, such as homeless adults (Smith, Meyers, & Delaney, 1998) and young adults (Smith, Godley, Godley, & Dennis, 2011). Systematic reviews have found the CRA to have among the most robust support for the treatment of AUDs (Miller & Wilbourne, 2002; Roozen et al., 2004).

The CRA has also been extended to facilitate treatment initiation, as an alternative to a confrontational Johnson Intervention. Specifically, a program has been developed that infuses principles of reinforcement learning into a training for families of individuals with AUDs geared toward motivating the target individual to initiate treatment, termed Community Reinforcement and Family Training (CRAFT; Meyers et al., 1998; Meyers & Smith, 1997). This approach has been supported by several clinical trials that have found CRAFT to increase treatment initiation at significantly higher rates than both Al Anon facilitation and Johnson Interventions (Dutcher et al., 2009; Miller et al., 1999). CRAFT engages approximately two-thirds of treatment-refusing drinkers into treatment (Meyers, Roozen, & Smith, 2011).

The second major form of reinforcement-based treatment is Contingency Management (CM). Unlike CRA, which focuses on the larger contingencies in a person's life, CM is microcosmic and seeks to reinforce protreatment outcomes directly using incentives (Stitzer & Petry, 2006). Specifically, CM provides vouchers for consumer goods or other positive reinforcers for individuals who are compliant with treatment, operationalized typically with treatment attendance and abstinence as measured via a drug test. What also distinguishes CM from the CRA is that it is very rapid in terms of responses. Changing the psychosocial contingencies surrounding drinking takes time and involves rewards that are often considerably

delayed, whereas CM provides immediate positive feedback on meeting a treatment goal.

There is extensive research support for CM for treating substance use disorders, including alcohol (Petry, Martin, Cooney, & Kranzler, 2000), opiates (McCaul, Stitzer, Bigelow, & Liebson, 1984), and stimulants (Higgins, Wong, Badger, Ogden, & Dantona, 2000). A meta-analysis of CM for substance use disorders found it to be consistently efficacious across 47 trials (Prendergast, Podus, Finney, Greenwell, & Roll, 2006). Furthermore, beyond addictive disorders, CM has been shown to enhance an array of treatment effects that apply to diverse health behaviors, including obesity (John et al., 2011; Volpp, John, et al., 2008), HIV risk behaviors (Ghitza, Epstein, & Preston, 2008) and anticoagulant medication compliance for individuals with heart disease (Volpp, Loewenstein, et al., 2008). With regard to alcohol, recent evidence shows that CM is associated with significant positive outcomes and that, with appropriate training, it can be implemented by community providers (Petry, Alessi, & Ledgerwood, 2012).

Of course, there is a challenge in implementing CM in AUD treatment because the typical assessment of use is breath alcohol content (BrAC), which is a state measure and reflects only current or very recent drug use. However, new biomarkers and technologies are circumventing this problem. For example, ethyl-glucuronide (EtG) is a biomarker of recent alcohol use that has a longer half-life and can be detected in urine for up to 2 days. A recent study found that EtG plus BrAC compliance tests substantially increased abstinence in a CM feasibility study (McDonell et al., 2012). Similarly, another study incorporated a Secure Continuous Remote Alcohol Monitoring (SCRAM) bracelet into a CM feasibility trial (Barnett, Tidey, Murphy, Swift, & Colby, 2011). The SCRAM device is a transdermal alcohol monitor that is used primarily in legal settings for monitoring alcohol use in adjudicated individuals. It provides around-the-clock monitoring and permitted verification of contingencies during a 2-week CM trial period, which significantly decreased drinking during this period (Barnett et al., 2011).

Cognitive-Behavioral Treatment

The most common form of contemporary cognitive-behavioral therapy (CBT) for AUDs has its foundations in social learning theory (Bandura, 1969), which broadly integrates principles of operant learning, observational learning, and cognitive psychology. A common feature of a CBT approach is an emphasis on alcohol as a powerful negative reinforcer (i.e., capable of alleviating aversive states). From this perspective, AUDs result from an overreliance on alcohol to manage negative affect due to an inability to use an adaptive, alternative means of removing aversive stimuli. More specifically, social learning theory proposes that substance-dependent individuals lack important coping skills for daily living and that such

skill deficits contribute significantly to substance dependence problems and relapse (Monti, Kadden, Rohsenow, Cooney, & Abrams, 2002).

A robust literature supports this etiological model: Individuals with AUDs have inferior coping skills compared with healthy controls; inferior coping skills to deal with high-risk circumstances over the course of treatment predict positive outcome; and low skill levels measured during treatment follow-up predict relapse (Miller, Westerberg, Harris, & Tonigan, 1996; Monti et al., 1990; for a review, see Monti et al., 2002). More indirect support for the approach comes from evidence that the treatment targets (e.g., managing negative emotional states) are associated with relapse (Marlatt & Gordon, 1985; Miller et al., 1996; Ramo & Brown, 2008).

The evidence of skills deficits has made skills training the centerpiece of CBT treatment. The skills targeted can be grouped into four primary domains: (1) interpersonal skills for building better relationships; (2) cognitive-emotional coping skills for mood regulation; (3) skills for improving daily living and dealing with stressful life events; and, (4) coping skills for managing environmental substance use cues (Monti et al., 2002). This approach has been manualized, but, even so, the treatment is highly idiographic and begins with an assessment of each patient's domains of vulnerability (Monti et al., 2002). These problem areas can be assessed via a number of instruments, including the Situational Competency Test (Chaney, O'Leary, & Marlatt, 1978), the Adaptive Skills Battery (Jones & Lanyon, 1981), and the Alcohol-Specific Role-Play Test (Abrams et al., 1991). Once vulnerabilities are systematically characterized, individuals work with the therapist on skill-building modules to address personal and general high-risk situations. These modules include drink refusal skills, development of sober supports, and conflict resolution skills (Monti et al., 2002).

There is strong empirical support for a CBT approach to AUDs. It has been found to be more efficacious than systematic desensitization, covert sensitization, aversion therapy (Hedberg & Campbell, 1974), traditional supportive therapy (Oei & Jackson, 1980, 1982), human relations training (Ferrell & Galassi, 1981), discussion groups (Chaney et al., 1978), mood management treatment (Monti et al., 1990), and group therapy (Eriksen, Björnstad, & Götestam, 1986), as well as no-treatment control conditions (Chaney et al., 1978). Although not all treatment outcome studies have provided unequivocally positive results, CBT has been shown to be of comparable efficacy to other active treatment conditions. For example, in Project MATCH, the CBT intervention was as effective as two other active treatments (Project MATCH Research Group, 1998).

One element of CBT that remains somewhat controversial is the fourth domain, that of coping with alcohol-related environmental cues. With foundations in classical conditioning, the strategy used in this domain is extinction to alcohol cues, that is, unreinforced exposure between environmental stimuli (e.g., an alcoholic beverage) and the associated behavioral response of drinking. It is referred to as cue exposure treatment (CET)

and uses exposure with response prevention in much the same way that it is used in the treatment of anxiety disorders (Foa & Kozak, 1986). A CET approach can be applied in a pure extinction model (e.g., Drummond & Glautier, 1994); however, a more common approach includes concurrent coping-skills training (Monti et al., 2001; Rohsenow et al., 2001). In the latter approach, the coping skills during the exposure pertain to both environmental cues in general and to craving that might derive from a number of different sources. In both cases, CET uses new learning to provide patients with more adaptive behavioral responses in the presence of drug cues.

Randomized controlled trials of CET for AUDs have revealed significant positive effects on an array of alcohol-related variables (Drummond & Glautier, 1994; Monti et al., 1993, 2001; Rohsenow et al., 2001; Sitharthan, Sitharthan, Hough, & Kavanagh, 1997). These effects appear to be the result of effects on craving and drink refusal self-efficacy (Monti et al., 2001). In contrast, however, a meta-analysis of CET for alcohol, tobacco, and opiates found limited support for the efficacy of CET (Conklin & Tiffany, 2002), making this approach somewhat controversial. When examined closely, however, differential results of treatment effects by drugs were present, with CET outcomes in trials for AUDs being uniformly positive and the opposite being the case for smoking cessation trials (MacKillop & Monti, 2007). In addition, the value of CET has been questioned in light of extensive evidence of context-dependent learning. In other words, the successful extinction and skills learning in treatment may not generalize to the patient's typical environmental context. Efforts have been made to improve the generalizability of extinction to alcohol cues (Hofmann, Huweler, MacKillop, & Kantak, 2012; MacKillop & Lisman, 2008; Stasiewicz, Brandon, & Bradizza, 2007), but these efforts have not yielded substantially greater extinction to date.

Relapse Prevention

The second common form of cognitive-behavioral treatment for AUDs is Relapse Prevention (RP; Marlatt & Gordon, 1985). A reality of treating AUDs is that many patients are not successful and return to drinking. Although precise definitions of relapse are contentious (Miller, 1996), estimates of the proportion of individuals who resume drinking are as high as 90% (Brandon, Vidrine, & Litvin, 2007). The RP approach evolved as an application of cognitive-behavioral approaches to the problem of relapse and as an alternative to disease-model notions of alcohol that emphasized endogenous cravings and withdrawal as the primary risk factors for relapse (for a review, see Hendershot, Witkiewitz, George, & Marlatt, 2011). More specifically, RP emphasizes the importance of environmental contextual factors (i.e., high-risk situations), cognitive processes (e.g., self-efficacy; expectancies), and the distinction between a single slip in the recovery

process (i.e., a lapse) and a complete return to pretreatment levels of problematic drinking (i.e., full relapse). A dichotomous (black-and-white) perspective approach to drinking is endemic of the traditional disease model, whereas a more nuanced continuum perspective is typical of a cognitive-behavioral approach. Similarly, RP explicitly addresses cognitive distortions, such as the abstinence violation effect, in which categorical thinking about a single drink can lead to shame, guilt, and the belief that a relapse is inevitable, leading to further drinking, thereby creating a vicious cycle.

Recognizing that long-term change will be challenging, the RP approach is a systematic intervention that is designed to prepare for possible lapses, minimize them to the extent possible, and, if a lapse takes place, to incorporate it into the recovery process in a positive way. Procedurally, RP begins with an idiographic assessment that delineates two basic components of risk for relapse: proximal determinants, or circumstances that can precipitate relapse, and potential covert antecedents, or less obvious circumstances that can trigger relapse (Larimer, Palmer, & Marlatt, 2003). Proximal determinants are variables that directly influence relapse, including such high-risk situations as social pressure or negative emotional states. In contrast, covert antecedents precede situations in which the individual is unprepared and thereby susceptible to relapse, such as driving by a favorite bar or spending time with certain people. Skills training is then used to prevent or otherwise cope with these circumstances.

A considerable literature supports the utility of RP for AUDs. A comprehensive review of 24 RP studies found that it was consistently superior to no intervention and equally effective as other active treatments, producing similar overall abstinence rates (Carroll, 1996). In addition, delayed emergence effects, or significant improvements found only at later follow-up points, were noted in a number of studies (Carroll, 1996). A meta-analysis of 26 studies using either RP alone or as an adjunct to other treatments found that it was effective in reducing alcohol use, and it was most effective when used as an adjunct to medication (Irvin, Bowers, Dunn, & Wang, 1999). McCrady (2000) reviewed the literature and concluded that there was robust support for RP in the treatment of AUDs and that it met the criteria for being an empirically supported treatment.

Motivational Interviewing

The preceding operant learning approaches emphasize the need to create and modify environmental contingencies that reinforce alternatives to drinking. The focus is on rearranging external aspects of the individual's environment, with relatively little focus on the individual's internal level of motivation. However, individuals with AUDs vary substantially in terms of motivation for change, and motivation is a critical factor in treatment engagement and response (Carney & Kivlahan, 1995; DiClemente & Hughes, 1990; Ryan, Plant, & O'Malley, 1995). Even high-quality

evidence-based treatments cannot be expected to help patients who have little motivation for change.

Motivational Interviewing (MI) is a treatment approach that directly addresses this issue and focuses on stimulating motivation to change. The foundations of the approach come from two distinct domains (Miller & Rose, 2009). The first is the humanistic tradition of Carl Rogers and its application to psychotherapy: namely, that, regardless of other active ingredients, a *sine qua non* of successful treatment is a supportive, nonjudgmental, and empathic interpersonal context (Rogers, 1959). In MI, an interpersonally and psychologically safe environment of therapy is intended to facilitate exploration of the pros and cons of personal change, free of outside judgment or coercion (Miller & Rollnick, 2012). The second domain is social psychological theories of cognitive dissonance (Festinger, 1957) and self-perception (Bem, 1972) that identify the innate motivational drive for individuals to find congruence between their perceptions of themselves and their actions. In the context of MI, treatment attempts to leverage the strength of individuals articulating aloud the incongruence in their life to facilitate positive change. This is the MI "one-two punch": an empathic therapeutic relationship for open discussion of alcohol use and selective emphasis of the associated discrepancies to catalyze change.

To achieve these goals, the therapeutic form of MI is based on four processes (Miller & Rollnick, 2012): engagement, focus on discrepancies, evocation of motivation, and change planning. First, the therapist *engages* the client by establishing an authentic and empathic working relationship using reflective listening and nonjudgmental discussion. This interactive process is intended to communicate acceptance and to be the antithesis of the confrontational or didactic stance that many problem drinkers have experienced from family members and professionals. It defines the client as the agent of change. In the second process, the therapist *focuses* the client by developing and maintaining an emphasis on change in the conversation. The third process is the most central: The therapist *evokes* the client's own motivations for change and facilitates the development of discrepancies between the client's drinking and his or her life goals. Following efforts to evoke motivation for change, the *planning process* comprises developing commitment to change and formulating a concrete plan of action.

Considerable evidence supports the efficacy of MI in a variety of treatment formats. Early trials yielded positive outcomes as a prelude to other active treatments (Bien, Miller, & Boroughs, 1993; Brown & Miller, 1993), and subsequent trials have supported it as a stand-alone intervention. For example, in the first multisite clinical trial, Project MATCH, a four-session MI intervention yielded equivalent outcomes to the two eight-session alternative interventions (Project MATCH Research Group, 1998). In addition, one of the MI matching hypotheses was supported, such that patients with high levels of trait anger had a better outcome in MI than in the other conditions (Project MATCH Research Group, 1998). In a second multisite

trial, MI significantly improved alcohol outcomes compared with a control intervention (Ball et al., 2007). Narrative reviews and meta-analyses have also provided consistent support for MI for AUDs (Hettema, Steele, & Miller, 2005; Miller & Wilbourne, 2002), and MI has been applied successfully to a wide diversity of health behaviors (Hettema et al., 2005).

Like most treatments, MI is associated with considerable variability in effects across studies and therapists (Ball et al., 2007; Hettema et al., 2005). This variability has led to efforts to understand the therapist characteristics and functional components of MI that moderate treatment outcomes. Two active components are broadly theorized to underlie the positive effects of MI: (1) a relational component that centers on empathy and the interpersonal relationship between the client and therapist, mapping onto its humanistic foundations; and (2) a technical component that focuses on the successful evocation and strengthening of "change talk" (i.e., client remarks favoring changing drinking and strategies for doing so) (Miller & Rose, 2009), mapping onto its social psychological foundations.

There is research support for both domains, which are neither incompatible nor mutually exclusive. Therapist empathy and interpersonal skills are positively related to MI outcome (Gaume, Gmel, & Daeppen, 2008; Moyers, Martin, Houck, Christopher, & Tonigan, 2009; Moyers, Miller, & Hendrickson, 2005). This finding converges with initial observations that empathic therapeutic relationships substantially predicted AUD treatment outcome (Miller & Baca, 1983; Miller, Taylor, & West, 1980), which was partially the inspiration for MI (Miller & Rose, 2009). Of particular interest, therapist–client interchanges coded using psycholinguistic analysis have provided support for the technical skill component as an active ingredient in MI outcome. Specifically, MI significantly decreases resistance and increases change talk (Amrhein, Miller, Yahne, Knupsky, & Hochstein, 2004; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004; Moyers & Martin, 2006; Moyers et al., 2007; Schoener, Madeja, Henderson, Ondersma, & Janisse, 2006). There is evidence of an inverse relationship between resistance and behavioral change: Greater resistance is associated with poorer outcome (Miller, Benefield, & Tonigan, 1993). Furthermore, change talk has been further fractionated into five domains—desire, ability, reasons, need, and commitment—and the lattermost aspect of change talk has been linked to long-term treatment response (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003). Importantly, this appears to be less related to absolute levels of commitment and more related to the *increasing trajectory* of commitment talk over the MI session (Amrhein et al., 2003). In a review of almost 20 studies, Apodaca and Longabaugh (2009) found consistent evidence for change talk, discrepancies, and therapist MI-consistent behavior as the mechanisms of change. Recent advances include using magnetoencephalography to track brain activity during MI sessions (Houck, Moyers, & Tesche, 2012). Taken together, this line of research represents the most extensive and rigorous attempt to unravel the mechanisms of action in

behavioral interventions for AUDs and, arguably, across all psychotherapy research.

Marital and Family Treatment

A common theme of the preceding treatments is their focus on the individual. In contrast, marital and family treatment (MFT) explicitly seeks to ameliorate alcohol problems through the medium of marital and family relationships. This approach emerged from early field and laboratory observations of the consequences of AUDs on an individual's interpersonal systems (e.g., Jacob & Seilhamer, 1987; Steinglass, 1981). Unsurprisingly, many alcoholics experience extensive marital and family problems (O'Farrell & Birchlery, 1987), and successful treatment is associated with better marital and family relationships (Moos, Finney, & Cronkite, 1990). From the MFT perspective, a person's spouse and family are often the most important people in an individual's life, which makes these relationships critical for success and failure. Furthermore, the relationship between AUDs and marital/family relationships is putatively bidirectional: AUDs contribute to marital and family discord, which recursively contribute to drinking itself and treatment prognosis (Maisto, O'Farrell, Connors, McKay, & Pelcovits, 1988; O'Farrell, 1995).

The goals of MFT with respect to AUDs are twofold. The first is to reduce excessive drinking, and the second is to reorganize the family environment to be more conducive to sobriety (O'Farrell, 1995). In the first case, treatment strategies include structured discussions about drinking, written behavioral contracts, and identification of family members' behaviors that instigate drinking. In the second case, MFT attempts to reengage the couple and the individual and the family in generally fulfilling relationships. This approach can be challenging because excessive drinking may result in enduring problems, both emotional and tangible (e.g., joblessness, debts). Moreover, although family relationships may contribute to drinking, this does not absolve patients of responsibility for their drinking and its consequences. Once drinking is managed successfully and relationship-based triggers and reinforcers for drinking have been addressed, MFT comprises efforts to enhance the quality of the marital relationship toward sustaining long-term success (Noel & McCrady, 1993; McCrady, Epstein, & Hirsch, 1999). This includes strategies such as increasing positive interchanges, problem solving, conflict resolution, planning shared recreational/leisure activities, training in communication skills, and behavior change agreements.

The benefits of MFT have been supported by a relatively large number of clinical studies. In an early study, couples receiving MFT and the medication disulfiram (described in greater detail below) improved significantly on various follow-up measures of marital adjustment (e.g., O'Farrell, Cutter, & Floyd, 1985), and these results were sustained at a 2-year follow-up

(O'Farrell, Cutter, Choquette, Floyd, & Bayog, 1992). Of note, however, no significant differences were found among groups in terms of reductions in drinking either at immediate or 2-year follow-up. A variant of MFT, Alcohol-Related Behavioral Couples Therapy (ABCT) (e.g., McCrady et al., 1999), has yielded consistently positive outcomes and comprises cognitive-behavioral treatment, spousal training to assist in changing antecedents of drinking, and a focus on improving the quality of the marital relationship. A recent study on ABCT revealed significant support for differences in positive long-term outcomes versus individual therapy (Walitzer & Dermen, 2004). Additionally, MFT efficacy was demonstrated in a study of women with alcohol use disorders and healthy male partners, with women receiving ABCT exhibiting better drinking outcomes over 6 months of treatment and 12 months of posttreatment follow-up compared with women in alcohol individual behavioral therapy (McCrady, Epstein, Cook, Jensen, & Hildebrandt, 2009). Notably, in one study of ABCT, relapse prevention components (e.g., AA encouragement, relapse prevention techniques) were found to offer little incremental benefit to improve long-term drinking outcomes over ABCT alone (McCrady, Epstein, & Kahler, 2004). In a systematic review of the literature, MFT was among the best supported interventions for AUDs (Miller, Wilbourne, & Hettema, 2003).

Brief Interventions

A last set of evidence-based treatment for AUDs is brief intervention of various kinds. Unlike the preceding psychological treatments that involve multisession regimens, these interventions typically use a single meeting to effect change. Although there is diversity among brief interventions, the most widely studied and most common format involves alcohol screening and, for individuals with a positive screen, a brief intervention and referral to more systematic treatment (referred to as SBIRT [screening, brief intervention, and referral to treatment]; Babor et al., 2007). This approach has its foundations in a public health model that seeks to minimize the burden of AUDs by addressing alcohol misuse in a wide array of diverse clinical settings where individuals have often not come to discuss alcohol misuse (Babor et al., 2007). These settings include emergency rooms, urgent care clinics, and primary care clinics. The public health model expands treatment throughout the health care system to provide a range of preventive/intervention strategies to all relevant individuals at risk for the condition. Put another way, this model asks: *Why wait for a person to experience severe alcohol problems and then seek treatment? Why not reach out to individuals, provide objective feedback about their drinking, and encourage them to seek treatment whenever possible?*

The foundations of brief interventions come from early studies revealing that a single session of objective feedback about relative drinking levels

and advice by a physician was as effective as more substantial interventions (Chafetz, 1961; Edwards et al., 1977; Kristenson, Ohlin, Hulten-Nosslin, Trell, & Hood, 1983). Although the SBIRT model includes referral to treatment, brief interventions have been shown to be efficacious independent of subsequent care in individual trials, systematic reviews, and meta-analyses (Babor et al., 2007; Bien, Miller, & Tonigan, 1993; Wilk, Jensen, & Havighurst, 1997). In addition, the brevity of this approach makes it highly cost-effective (e.g., Fleming et al., 2002). Importantly, brief interventions are not a panacea, but rather a highly effective first level of care. This is especially the case for individuals with less severe alcohol problems (Moyer, Finney, Swearingen, & Vergun, 2002). More treatment has been associated with better outcomes with more severe patients (Berglund et al., 2003), and, for these individuals, here SBIRT referral to treatment may be most valuable. Even so, SBIRT interventions are the strategy of first resort and are recommended as standard practice in diverse clinical settings and, ultimately, as population wide.

Efficacious Medications for Treating AUDs

Contemporary AUD treatment typically takes place in a multidisciplinary context. Clinical psychologists are often part of a treatment team that includes physicians, nurses, social workers, and other staff playing diverse roles. As there are currently four medications approved by the Food and Drug Administration (FDA) for treating AUDs, clinical psychologists are likely to see patients receiving these or related medications. The approved medications are disulfiram (trade name Antabuse), naltrexone in both tablet form (trade name ReVia) and depot injection (trade name Vivitrol), and acamprosate (trade name Campra). Alternatively, in individual treatment settings, such as a private practice, psychologists may collaborate with a physician to augment psychological interventions with these medications. Based on the increasing relevance of pharmacotherapies for clinical psychologists, in this section we review the efficacy and putative mechanisms of action for each medication.

Disulfiram was the first medication to be approved for treating AUDs, in 1951, and it acts by interfering with the metabolism of alcohol. If a person drinks while taking disulfiram, a buildup of acetaldehyde accumulates in the bloodstream, and he or she experiences an unpleasant syndrome, including nausea, dizziness, rapid heart rate, headache, and facial flushing. This is the same as the flushing syndrome that is present in individuals who have a genetic variant resulting in deficient alcohol metabolism and the same buildup of acetaldehyde. Based on this mutation, these individuals, typically of East Asian ancestry, are relatively protected against developing alcohol problems (Luczak, Wall, Cook, Shea, & Carr, 2004). By inducing these symptoms, the ostensible mechanism of disulfiram is a combination of punishment and aversive conditioning: The person elects not to drink to

avoid the punishing effects of flushing syndrome and, if they do drink, the experience becomes associated with an intensely aversive experience.

A recent meta-analysis of 11 RCTs supported the efficacy of disulfiram (Jørgensen, Pedersen, & Tønnesen, 2011), but it is notable that a number of earlier trials yielded inconsistent findings (Suh, Pettinati, Kampman, & O'Brien, 2006). This is in part because issues of patient compliance consistently arise with disulfiram. When patients face the prospect of taking disulfiram and inducing negative effects when they drink alcohol, they frequently elect simply to stop taking disulfiram, rather than stop drinking. For example, in a large multisite RCT, only 20% of the 577 completed patients were deemed compliant (Fuller et al., 1986). Efforts to improve compliance have focused increasingly on disulfiram administered in a supervised fashion (e.g., Azrin, Sisson, Meyers, & Godley, 1982; Chick et al., 1992), and these studies have yielded significantly better outcomes (Krampe & Ehrenreich, 2010). One such example is the marital treatment program described by O'Farrell (1995), which uses spousal monitoring of disulfiram administration to maintain compliance and engage the spouse in treatment. For the alcoholic, compliance demonstrates a show of good faith and commitment to recovery. This spousal dimension also illustrates the more subtle psychological mechanisms of disulfiram, including a therapeutic ritual, frequent renewal of commitment, and ongoing reinforcement of sobriety (Krampe & Ehrenreich, 2010).

Naltrexone, both in tablet and injectable form, is an endogenous opioid antagonist that putatively reduces alcohol motivation by reducing its rewarding effects in the brain (King, Volpicelli, Frazer, & O'Brien, 1997; McCaul, Wand, Eissenberg, Rohde, & Cheskin, 2000; Swift, Whelihan, Kuznetsov, & Buongiorno, 1994) and potentially attenuating alcohol cue reactivity (Gastfriend, 2011; Rohsenow et al., 2000). In general, positive naltrexone effects have been consistent across studies but relatively small in magnitude and restricted to heavy drinking (Pettinati et al., 2006). In a meta-analysis, oral naltrexone was associated with an approximately 30% lower rate of heavy drinking compared with placebo but had no significant effect on abstinence (Pettinati et al., 2006). However, adherence is a significant issue with oral naltrexone as well (Chick et al., 2000; Volpicelli et al., 1997), which was the basis for developing a long-acting (monthly) intramuscular depot formulation. Injectable naltrexone has been shown to be efficacious compared with placebo (Garbutt et al., 2005; O'Malley, Garbutt, Gastfriend, Dong, & Kranzler, 2007). Recent subgroup analyses suggest that it significantly improves outcomes for severely dependent individuals (Pettinati et al., 2011).

The most recently FDA-approved medication is acamprosate, which modulates glutamate neurotransmission and putatively ameliorates neural hyperexcitability induced by protracted withdrawal (Kiefer & Mann, 2010; Mason & Heyser, 2010). Unlike disulfiram or naltrexone, acamprosate is not intended to affect alcohol's acute effects, but to alleviate the

consequences of chronic drinking. Acamprosate has been supported in a number of placebo-controlled trials (Rösner et al., 2010), but two recent large-scale trials in the United States did not find significant positive effects (Anton et al., 2006; Berger Fisher, Brondino, Bohn, Gwyther, et al., 2012). Importantly, although the mechanism of acamprosate is thought to be the restoration of the appropriate balance of excitatory and inhibitory activity, there is limited evidence for this hypothesis in humans (Kiefer & Mann, 2010). More generally, the efficacy of these medications is substantially better understood than their pharmacological and psychological mechanisms of action.

CONCLUSIONS

The goal of this chapter was to review both controversial and evidence-based approaches to the treatment of AUDs. This volume and its first edition (Lilienfeld, Lynn, & Lohr, 2003) attest to the long history of controversial treatments in clinical psychology, but the controversies in this area seem particularly acrimonious. This is in large part because of the two categories of clinicians in AUD treatment: those who have a strong personal stake in a specific approach and those who want to apply the best strategies available that are supported by research. Returning to the original distinction, the common theme underlying all four controversial treatments is that they reflect a disjunction between craftsmen–clinicians and scientist–clinicians.

However, there are important distinctions among the controversial treatments. Two cases, Johnson Interventions and DARE, are fairly clear instances of pseudoscientific approaches that are both widespread and persistent, despite negligible or negative evidence. At the other end of the spectrum is controlled drinking, where one potential treatment outcome is commonly proscribed despite evidence of its viability for some individuals. Between the two is AA, which in some ways fits the mold of a pseudoscientific approach, but has been increasingly studied by behavioral scientists. These recent studies provide support for AA's utility, although in its capacity as a mutual-help group and for certain individuals, not as a universal intervention. There is no question that AA's etiological approach, more or less unchanged since its inception, is outmoded and that its treatment prescription is dogmatic and unconditional. We would argue that formal treatment facilities that adopt the AA approach entirely and imbue it throughout their programming, as in the "Minnesota Model," are misguided and are taking both the "baby" and the "bathwater." However, in contrast to the Johnson Intervention and DARE, research evidence exists that demonstrates that AA contains active ingredients that help people change their drinking, albeit these are not AA-specific so much as mutual-help group-specific.

The review of the evidence-based approaches to AUDs reveals the wide diversity of options available. One common theme is that several reinforcement-based approaches and cognitive-behavioral therapy emerged from learning theory as the application of basic principles to clinical problems. They can be thought of as first-wave and second-wave interventions, respectively (Hayes, 2004). However, this is not the case for all the evidence-based treatments: Motivational interviewing combines humanistic and social psychological approaches, brief interventions apply a public health model, and pharmacotherapies apply a traditional medical model. There are commonalities across the evidence-based practices in terms of the psychological treatment regimens. All include idiographic assessment and customization of the principles of the approach to the individual. They commonly avoid a one-size-fits-all approach. The other consistent theme across these approaches is their emphasis on scientific evidence as essential in the AUD treatment enterprise.

In light of these various evidence-based options, the most cogent recommendation remains Hester and Miller's (2003) call for *informed eclecticism*—implementing a diversity of treatment strategies that have met empirical benchmarks in clinical practice. There is no silver bullet for treating AUDs, but there is an armamentarium of tools and approaches that have a strong empirical basis and are the most appropriate for clinical practice. A good example of informed eclecticism was the manualized combined behavioral intervention (CBI) in the recent multisite clinical trial named Project COMBINE (Anton et al., 2006). This major initiative by NIAAA examined the prospect of combining naltrexone with acamprosate in the treatment of AUDs and included a combination of empirically supported elements of psychological interventions, including aspects of MI, CBT, and RP. Study clinicians followed a protocol that provided both structure and the opportunity to implement the modules that were most appropriate for a given patient. This flexibility demonstrates that the evidence-based strategies identified are by no means incompatible or mutually exclusive with each other. Indeed, there is some evidence that MI is synergistic with other active treatments (Hettema et al., 2005) and specific combinations of these treatments might ultimately become recommended practices.

Although some of the controversial treatments without demonstrated efficacy discussed in this chapter remain widespread, there is evidence for a gradual, albeit too slow, diffusion of evidence-based practices into alcohol treatment. However, a shared value among the constituencies is to achieve the best possible outcomes for patients. Where nonevidence-based or pseudoscientific approaches are routinely used, it is not for perverse or subversive reasons, but because the clinicians involved genuinely believe that these approaches are the most useful based on their experience or intuition. As such, there is little motivation to change. Therefore, we contend that the onus continues to be on clinical scientists to persevere toward making the

dominant clinical approach one that is evidence-based. It is not a matter of "if you build it, they will come." Although the field has developed robustly supported approaches, the adoption problem remains. Thus, the dissemination model also needs to change, with clinical scientists actively engaging in ways to systematically and skillfully transmit best-practices into the field. This may come from top-down strategies, such as including training in evidence-based practices as a requirement for clinician credentialing, or from bottom-up strategies, such as developing easily accessible treatment manuals and providing training at clinical conferences. In either case, the landscape of alcohol treatment will not change appreciably without an active scientific constituency seeking to make that happen.

GLOSSARY

Alcoholics Anonymous (AA): Founded in 1935, AA is a nonprofit, grass-roots lay treatment for alcoholism. Core tenets include a chronic disease model of alcoholism, treatable only by total abstinence; adherence to its 12-step ideology, including an emphasis on nondenominational spirituality; and the use of fellowship among alcoholics to maintain abstinence.

Community-reinforcement approach (CRA): A psychotherapeutic intervention based on operant conditioning that attempts to reduce drinking behavior by restructuring the individual's reinforcement contingencies to increase the value of nondrinking behaviors.

Coping/social skills training (CSST): A psychotherapeutic form of treatment for alcohol dependence from the perspective of social learning theory. CSST attempts to reduce drinking by improving individuals' ability to cope with relationships, events in their lives, and their own negative moods, often in the context of alcohol cues.

Cue exposure treatment (CET): Both a component of CSST and a stand-alone treatment, CET is based on a classical conditioning model of alcoholism and leads to drinking reduction or abstinence by exposing the alcoholic to alcohol-related cues.

Disulfiram (commercial name: Antabuse): The most common antidrinking medication, disulfiram induces flushing, rapid or irregular heartbeat, sweating, nausea, and dizziness by preventing the release of aldehyde dehydrogenase, a liver enzyme required to break down alcohol.

Marital and family therapy (MFT): A psychotherapeutic treatment that uses the marital relationship as the forum for reducing drinking behavior. Clinically, couples may learn improved communication skills, develop problem solving for conflict resolution, and address intimacy issues. One premise of these efforts is that improvement in the marital relationship will decrease the appeal of alcohol. Another is that spousal engagement will help problem drinkers to develop and maintain their treatment goals.

Motivational enhancement therapy (MET): A clinical intervention that identifies discrete stages of change and attempts to facilitate the client's progress through these stages toward abstinence or a reduction in drinking.

Relapse prevention (RP): A form of psychotherapeutic intervention based on social learning theory that focuses on incorporating plans for probable relapse into the

treatment process. RP uses skills training and cognitive techniques to reduce the likelihood and severity of relapse.

Social learning theory (SLT): A broad theoretical approach to human behavior that incorporates principles of classical and operant conditioning, cognitive psychology, and observational learning. In applications to drinking problems, SLT provides the basis for strategies that include coping skills, modeling, rehearsal, attribution, and enhancement of self-efficacy.

ACKNOWLEDGMENTS

Preparation of this chapter was partially supported by National Institutes of Health Grant K23 AA016936 (to James MacKillop). We are grateful for the original insights by the authors of this chapter in the previous edition of this volume: Drs. Stephen Lisman, Allison Weinstein, and Deborah Rosenbaum.

REFERENCES

A&E. (2012). Retrieved December 7, 2012, from *www.aetv.com/intervention/about*.

Abrams, D. B., Binkoff, J. A., Zwick, W. R., Liepman, M. R., Nirenberg, T. D., Munroe, S. M., et al. (1991). Alcohol abusers' and social drinkers' responses to alcohol-relevant and general situations. *Journal of Studies on Alcohol and Drugs, 52*(5), 409–414.

Aharonovich, E., Amrhein, P. C., Bisaga, A., Nunes, E. V., & Hasin, D. S. (2008). Cognition, commitment language, and behavioral change among cocaine-dependent patients. *Psychology of Addictive Behaviors, 22*(4), 557–562.

Alcoholics Anonymous. (2012). Service material from the General Service Office.

Alford, G. S. (1980). Alcoholics Anonymous: An empirical outcome study. *Addictive Behaviors, 5*(4), 359–370.

Amrhein, P. C., Miller, W. R., Yahne, C., Knupsky, A., & Hochstein, D. (2004). Strength of client commitment language improves with therapist training in motivational interviewing. *Alcoholism: Clinical and Experimental Research, 28*(5).

Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology, 71*(5), 862.

Anton, R. F., O'Malley, S. S., Ciraulo, D. A., Cisler, R. A., Couper, D., Donovan, D. M., et al. (2006). Combined pharmacotherapies and behavioral interventions for alcohol dependence. *Journal of the American Medical Society, 295*(17), 2003–2017.

Apodaca, T. R., & Longabaugh, R. (2009). Mechanisms of change in motivational interviewing: A review and preliminary evaluation of the evidence. *Addiction, 104*(5), 705–715.

Armor, D. J., Polich, J. M., & Stambul, H. B. (1976). *Alcoholism and treatment*. Santa Monica, CA: Rand Corporation.

Assistance in Recovery. (2012). Fact sheet. Available at *http://a-i-r.com/air/pressroom/Fact%20Sheet.pdf*.

Azrin, N. H. (1976). Improvements in the community-reinforcement approach to alcoholism. *Behaviour Research and Therapy, 14*(5), 339–348.

Azrin, N. H., Sisson, R. W., Meyers, R., & Godley, M. (1982). Alcoholism treatment by disulfiram and community reinforcement therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 13*(2), 105–112.

Babor, T. F., McRee, B. G., Kassebaum, P. A., Grimaldi, P. L., Ahmed, K., & Bray, J. (2007). Screening, brief intervention, and referral to treatment (SBIRT): Toward a public health approach to the management of substance abuse. *Focus, 28*(3), 7–30.

Ball, S. A., Martino, S., Nich, C., Frankforter, T. L., Van Horn, D., Crits-Christoph, P., et al. (2007). Site matters: Multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *Journal of Consulting and Clinical Psychology, 75*(4), 556–567.

Bandura, A. (1969). *Principles of behavior modification.* New York: Holt, Rinehart & Winston.

Barnett, N. P., Tidey, J., Murphy, J. G., Swift, R., & Colby, S. M. (2011). Contingency management for alcohol use reduction: A pilot study using a transdermal alcohol sensor. *Drug and Alcohol Dependence, 118*(2), 391–399.

Bem, D. J. (1972). *Advances in experimental social psychology:* Vol. 6. *Self-Perception Theory.* New York: Academic Press.

Berger, L., Fisher, M., Brondino, M., Bohn, M., Gwyther, R., Longo, L., et al. (2012). Efficacy of acamprosate for alcohol dependence in a family medicine setting in the United States: A randomized, double-blind, placebo-controlled study. *Alcoholism: Clinical and Experimental Research, 37*(4), 668–672.

Berglund, M., Thelander, S., Salaspuro, M., Franck, J., Andreasson, S., & Ojehagen, A. (2003). Treatment of alcohol abuse: An evidence-based review. *Alcohol Clinical and Experimental Research, 27*(10), 1645–1656.

Bien, T. H., Miller, W. R., & Boroughs, J. M. (1993). Motivational interviewing with alcohol outpatients. *Behavioural and Cognitive Psychotherapy, 21*(4), 347–356.

Bien, T. H., Miller, W. R., & Tonigan, J. S. (1993). Brief interventions for alcohol problems: A review. *Addiction, 88*(3), 315–336.

Bigelow, G. (2001). An operant behavioral perspective on alcohol abuse and dependence. In N. Healther, T. J. Peters, & T. Stockwell (Eds.), *International handbook of alcohol dependence and problems* (pp. 299–315). Hoboken, NJ: Wiley.

Brandon, T. H., Vidrine, J. I., & Litvin, E. B. (2007). Relapse and relapse prevention. *Annual Review of Clinical Psychology, 3*, 257–284.

Brown, J. M., & Miller, W. R. (1993). Impact of motivational interviewing on participation and outcome in residential alcoholism treatment. *Psychology of Addictive Behaviors, 7*(4), 211–218.

Brown, T. G., Seraganian, P., Tremblay, J., & Annis, H. (2002). Matching substance abuse aftercare treatments to client characteristics. *Addictive Behavior, 27*(4), 585–604.

Bufe, C. (1991). *Alcoholics Anonymous: Cult or cure?* (2nd ed.). Tuscon, AZ: See Sharp Press.

Carney, M. M., & Kivlahan, D. R. (1995). Motivational subtypes among veterans seeking substance abuse treatment: Profiles based on stages of change. *Psychology of Addictive Behaviors, 9*(2), 135–142.

Carroll, K. M. (1996). Relapse prevention as a psychosocial treatment: A review

of controlled clinical trials. *Experimental and Clinical Psychopharmacology, 4*(1), 46–54.
Chafetz, M. E. (1961). A procedure for establishing therapeutic contact with the alcoholic. *Quarterly Journal of Studies on Alcohol, 22*, 325–328.
Chaney, E. F., O'Leary, M. R., & Marlatt, G. A. (1978). Skill training with alcoholics. *Journal of Consulting and Clinical Psychology, 46*(5), 1092–1104.
Chiauzzi, E. J., & Liljegren, S. (1993). Taboo topics in addiction treatment: An empirical review of clinical folklore. *Journal of Substance Abuse Treatment, 10*(3), 303–316.
Chick, J., Anton, R., Checinski, K., Croop, R., Drummond, D. C., Farmer, R., et al. (2000). A multicentre, randomized, double-blind, placebo-controlled trial of naltrexone in the treatment of alcohol dependence or abuse. *Alcohol and Alcoholism, 35*(6), 587–593.
Chick, J., Gough, K., Falkowski, W., Kershaw, P., Hore, B., Mehta, B., et al. (1992). Disulfiram treatment of alcoholism. *British Journal of Psychiatry, 161*(1), 84–89.
Clayton, R., Leukefeld, C., Harrington, N., & Cattarello, A. (1996). DARE (Drug Abuse Resistance Education): Very popular but not very effective. In C. B. McCoy, L. R. Metsch, & J. A. Inciardi (Eds.), *Intervening with drug-involved youth* (pp. 101–109). Thousand Oaks, CA: Sage.
Conklin, C. A., & Tiffany, S. T. (2002). Applying extinction research and theory to cue-exposure addiction treatments. *Addiction, 97*(2), 155.
Coombs, R. H., & Ziedonis, D. (1995). *Handbook on drug abuse prevention: A comprehensive strategy to prevent the abuse of alcohol and other drugs*. Boston: Allyn & Bacon.
Curtis, S. L., & Eby, L. T. (2010). Recovery at work: The relationship between social identity and commitment among substance abuse counselors. *Journal of Substance Abuse Treatment, 39*(3), 248–254.
Davies, D. L. (1962). Normal drinking in recovered addicts. *Quarterly Journal of Studies on Alcohol, 23,* 94–104.
Davis, A. K., & Rosenberg, H. (2012). Acceptance of non-abstinence goals by addiction professionals in the United States. *Psychology of Addictive Behaviors*, *27*, 1102–1109.
DiClemente, C. C., & Hughes, S. O. (1990). Stages of change profiles in outpatient alcoholism treatment. *Journal of Substance Abuse, 2*(2), 217–235.
Ditman, K. S., Crawford, C. G., Forgy, E. W., Moskowitz, H., & MacAndrew, C. (1967). A controlled experiment on the use of court probation for drunk arrests. *American Journal of Psychiatry, 124,* 160–163.
Donnermeyer, J. F. (2000). Parents' perceptions of a school-based prevention education program. *Journal of Drug Education, 30*(3), 325–342.
Donnermeyer, J. F., & Wurschmidt, T. N. (1997). Educators' perceptions of the DARE program. *Journal of Drug Education, 27*(3), 259–276.
Drug Abuse Resistance Education. (2009). *D.A.R.E. 2009 annual report: Empowering children to lead safe and healthy lives*. Los Angeles: Author.
Drummond, D. C., & Glautier, S. (1994). A controlled trial of cue exposure treatment in alcohol dependence. *Journal of Consulting and Clinical Psychology, 62*(4), 809.
Dukes, R. L., Ullman, J. B., & Stein, J. A. (1996). Three-year follow-up of drug abuse resistance education (DARE). *Evaluation Review, 20*(1), 49–66.

Dutcher, L., Anderson, R., Moore, M., Luna-Anderson, C., Meyers, R., Delaney, H. D., et al. (2009). Community reinforcement and family training (CRAFT): An effectiveness study. *Journal of Behavior Analysis in Health, Sports, Fitness and Medicine, 2*, 80–93.

Edwards, G., Orford, J., Egert, S., Guthrie, S., Hawker, A., Hensman, C., et al. (1977). Alcoholism: A controlled trial of "treatment" and "advice." *Journal of Studies on Alcohol and Drugs, 38*(5), 1004.

Ellis, A., & Schoenfeld, E. (1990). Divine intervention and the treatment of chemical dependency. *Journal of Substance Abuse, 2*, 459–469.

Ennett, S. T., Tobler, N. S., Ringwalt, C. L., & Flewelling, R. L. (1994). How effective is drug abuse resistance education?: A meta-analysis of Project DARE outcome evaluations. *American Journal of Public Health, 84*(9), 1394–1401.

Eriksen, L., Björnstad, S., & Götestam, K. G. (1986). Social skills training in groups for alcoholics: One-year treatment outcome for groups and individuals. *Addictive Behaviors, 11*(3), 309–329.

Eysenck, H. J. (1994). The outcome problem in psychotherapy: What have we learned? *Behaviour Research and Therapy, 32*(5), 477–495.

Ferrell, W. L., & Galassi, J. P. (1981). Assertion training and human relations training in the treatment of chronic alcoholics. *Substance Use and Misuse, 16*(5), 959–968.

Ferri, M., Amato, L., & Davoli, M. (2006). Alcoholics Anonymous and other 12-step programmes for alcohol dependence. *Cochrane Database of Systematic Reviews, 3*(2).

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press

Fleming, M. F., Mundt, M. P., French, M. T., Manwell, L. B., Stauffacher, E. A., & Barry, K. L. (2002). Brief physician advice for problem drinkers: Long-term efficacy and benefit-cost analysis. *Alcohol Clinical and Experimental Research, 26*(1), 36–43.

Fletcher, A. M. (2002). *Sober for good: New solutions for drinking problems—advice from those who have succeeded*. Boston: Houghton Mifflin Harcourt.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin, 99*(1), 20–35.

Fuller, R. K., Branchey, L., Brightwell, D. R., Derman, R. M., Emrick, C. D., Iber, F. L., et al. (1986). Disulfiram treatment of alcoholism. *Journal of the American Medical Association, 256*(11), 1449–1455.

Galanter, M. (1993). *Network therapy for alcohol and drug abuse*. New York: Basic Books.

Garbutt, J. C., Kranzler, H. R., O'Malley, S. S., Gastfriend, D. R., Pettinati, H. M., Silverman, B. L., et al. (2005). Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence. *Journal of the American Medical Association, 293*(13), 1617–1625.

Garrett, J., Landau, J., Shea, R., Stanton, M. D., Baciewicz, G., & Brinkman-Sull, D. (1998). The ARISE Intervention: Using family and network links to engage addicted persons in treatment. *Journal of Substance Abuse Treatment, 15*(4), 333–343.

Gastfriend, D. R. (2011). Intramuscular extended-release naltrexone: Current evidence. *Annals of the New York Academy of Sciences, 1216*(1), 144–166.

Gaume, J., Gmel, G., & Daeppen, J. B. (2008). Brief alcohol interventions: Do

counsellors' and patients' communication characteristics predict change? *Alcohol and Alcoholism, 43*(1), 62–69.

Ghitza, U. E., Epstein, D. H., & Preston, K. L. (2008). Contingency management reduces injection-related HIV risk behaviors in heroin and cocaine using outpatients. *Addictive Behaviors, 33*(4), 593–604.

Groh, D. R., Jason, L. A., & Keys, C. B. (2008). Social network variables in alcoholics anonymous: A literature review. *Clinical Psychology Review, 28*(3), 430–450.

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy, 35*(4), 639–665.

Heather, N., Adamson, S. J., Raistrick, D., & Slegg, G. P. (2010). Initial preference for drinking goal in the treatment of alcohol problems: I. Baseline differences between abstinence and non-abstinence groups. *Alcohol and Alcoholism, 45*(2), 128–135.

Heather, N., & Dawe, S. (2005). Level of impaired control predicts outcome of moderation-oriented treatment for alcohol problems. *Addiction, 100*(7), 945–952.

Hedberg, A. G., & Campbell, L. (1974). A comparison of four behavioral treatments of alcoholism. *Journal of Behavior Therapy and Experimental Psychiatry, 5*(3), 251–256.

Hendershot, C. S., Witkiewitz, K., George, W. H., & Marlatt, G. A. (2011). Relapse prevention for addictive behaviors. *Substance Abuse Treatment, Prevention, and Policy, 6*(1), 17.

Hester, R. K., & Delaney, H. D. (1997). Behavioral self-control program for Windows: Results of a controlled clinical trial. *Journal of Consulting and Clinical Psychology, 65*(4), 686–693.

Hester, R. K., Delaney, H. D., & Campbell, W. (2011). *ModerateDrinking.com* and moderation management: Outcomes of a randomized clinical trial with non-dependent problem drinkers. *Journal of Consulting and Clinical Psychology, 79*(2), 215–224.

Hester, R. K., & Miller, W. R. (2003). *Handbook of alcoholism treatment approaches: Effective alternatives*. Boston: Allyn & Bacon.

Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annual Review of Clinical Psychology, 1*, 91–111.

Higgins, S. T., Heil, S. H., & Lussier, J. P. (2004). Clinical implications of reinforcement as a determinant of substance use disorders. *Annual Review of Clinical Psychology, 55*, 431–461.

Higgins, S. T., Wong, C. J., Badger, G. J., Ogden, D. E., & Dantona, R. L. (2000). Contingent reinforcement increases cocaine abstinence during outpatient treatment and 1 year of follow-up. *Journal of Consulting and Clinical Psychology, 68*(1), 64–72.

Hodgins, D. C., Ching, L. E., & McEwen, J. (2009). Strength of commitment language in motivational interviewing and gambling outcomes. *Psychology of Addictive Behaviors, 23*(1), 122–130.

Hofmann, S. G., Huweler, R., MacKillop, J., & Kantak, K. M. (2012). Effects of D-cycloserine on craving to alcohol cues in problem drinkers: Preliminary findings. *American Journal of Drug and Alcohol Abuse, 38*(1), 101–107.

Houck, J. M., Moyers, T. B., & Tesche, C. D. (2012). Through a glass darkly: Some

insights on change talk via magnetoencephalography. *Psychology of Addictive Behaviors, 27,* 489–500.

Humphreys, K., & Moos, R. (2001). Can encouraging substance abuse patients to participate in self-help groups reduce demand for health care?: A quasi-experimental study. *Alcoholism: Clinical and Experimental Research, 25*(5), 711–716.

Humphreys, K., & Moos, R. H. (2007). Encouraging posttreatment self-help group involvement to reduce demand for continuing care services: Two-year clinical and utilization outcomes. *Alcoholism: Clinical and Experimental Research, 31*(1), 64–68.

Hunt, G. M., & Azrin, N. H. (1973). A community-reinforcement approach to alcoholism. *Behavior Research and Therapy, 11*, 91–104.

Irvin, J. E., Bowers, C. A., Dunn, M. E., & Wang, M. C. (1999). Efficacy of relapse prevention: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 67*(4), 563.

Jacob, T., & Seilhamer, R. A. (1987). *Alcoholism and family interaction*. New York: Plenum Press.

Jellinek, E. M. (1946). Phases in the drinking history of alcoholics. *Quarterly Journal of Studies on Alcohol, 7*(1), 1–88.

Jellinek, E. M. (1960). *The disease concept of alcoholism*. New Haven, CT: Hillhouse Press.

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., & Volpp, K. G. (2011). Financial incentives for extended weight loss: A randomized, controlled trial. *Journal of General Internal Medicine, 26*(6), 621–626.

Johnson, V. E. (1986). *Intervention: How to help someone who doesn't want help: A step-by-step guide for families and friends of chemically dependent persons*. Center City, MN: Hazelden.

Jones, S. L., & Lanyon, R. I. (1981). Relationship between adaptive skills and outcome of alcoholism treatment. *Journal of Studies on Alcohol and Drugs, 42*(5), 521.

Jørgensen, C. H., Pedersen, B., & Tønnesen, H. (2011). The efficacy of disulfiram for the treatment of alcohol use disorder. *Alcoholism: Clinical and Experimental Research, 35*(10), 1749–1758.

Kalb, M., & Propper, M. S. (1976). The future of alcohology: Craft or science? *American Journal of Psychiatry, 133*(6), 641–645.

Kaskutas, L. A. (2009). Alcoholics anonymous effectiveness: Faith meets science [Review]. *Journal of Addictive Diseases, 28*(2), 145–157.

Kelly, J. F. (2013). Mutual-help groups. In B. S. McCrady & E. E. Epstein (Eds.), *Addictions: A comprehensive guidebook*. New York: Oxford University Press.

Kelly, J. F., Hoeppner, B., Stout, R. L., & Pagano, M. (2012). Determining the relative importance of the mechanisms of behavior change within Alcoholics Anonymous: A multiple mediator analysis. *Addiction, 107*(2), 289–299.

Kelly, J. F., Magill, M., & Stout, R. L. (2009). How do people recover from alcohol dependence?: A systematic review of the research on mechanisms of behavior change in Alcoholics Anonymous. *Addiction Research and Theory, 17*(3), 236–259.

Kelly, J. F., & Moos, R. (2003). Dropout from 12-step self-help groups: Prevalence, predictors, and counteracting treatment influences. *Journal of Substance Abuse Treatment, 24*(3), 241–250.

Kiefer, F., & Mann, K. (2010). Acamprosate: How, where, and for whom does it work?: Mechanism of action, treatment targets, and individualized therapy. *Current Pharmaceutical Design, 16*(19), 2098–2102.

King, A. C., Volpicelli, J. R., Frazer, A., & O'Brien, C. P. (1997). Effect of naltrexone on subjective alcohol response in subjects at high and low risk for future alcohol dependence. *Psychopharmacology, 129*(1), 15–22.

Koch, K. A. (1994). The D.A.R.E. (Drug Abuse Resistance Education) program. In J. A. Lewis (Ed.), *Addictions: Concepts and strategies for treatment* (pp. 359–364). Gaithersburg, MD: Aspen .

Kownacki, R. J., & Shadish, W. R. (1999). Does Alcoholics Anonymous work?: The results from a meta-analysis of controlled experiments. *Substance Use and Misuse, 34*(13), 1897–1916.

Krampe, H., & Ehrenreich, H. (2010). Supervised disulfiram as adjunct to psychotherapy in alcoholism treatment. *Current Pharmaceutical Design, 16*(19), 2076–2090.

Kristenson, H., Ohlin, H., Hulten-Nosslin, M. B., Trell, E., & Hood, B. (1983). Identification and intervention of heavy drinking in middle-aged men: Results and follow-up of 24–60 months of long-term study with randomized controls. *Alcoholism: Clinical and Experimental Research, 7*(2), 203–209.

Landau, J., Stanton, M. D., Brinkman-Sull, D., Ikle, D., McCormick, D., Garrett, J., et al. (2004). Outcomes with the ARISE approach to engaging reluctant drug-and alcohol-dependent individuals in treatment. *American Journal of Drug and Alcohol Abuse*, *30*(4), 711–748.

Larimer, M. E., Palmer, R. S., & Marlatt, A. G. (2003). Relapse prevention: An overview of Marlatt's cognitive-behavioral model (Vol. 3). In E. F. McCance-Katz, & H. W. Clark (Eds.), *Psychosocial treatments* (pp. 1–18). New York: Brunner-Routledge.

Levine, H. G. (1978). The discovery of addiction. Changing conceptions of habitual drunkenness in America. *Journal of Studies on Alcohol, 39*(1), 143–174.

Liepman, M. R., Nirenberg, T. D., & Begin, A. M. (1989). Evaluation of a program designed to help family and significant others to motivate resistant alcoholics into recovery. *American Journal of Drug and Alcohol Abuse*, *15*(2), 209–221.

Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (Eds.). (2003). *Science and pseudoscience in clinical psychology*. New York: Guilford Press.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*(1), 53–70.

Lisman, S. A. (1974). Alcoholic "blackout": State dependent learning? *Archives of General Psychiatry, 30*(1), 46–53.

Loneck, B., Garrett, J. A., & Banks, S. M. (1996a). A comparison of the Johnson intervention with four other methods of referral to outpatient treatment. *American Journal of Drug and Alcohol Abuse, 22*(2), 233–246.

Loneck, B., Garrett, J. A., & Banks, S. M. (1996b). The Johnson intervention and relapse during outpatient treatment. *American Journal of Drug and Alcohol Abuse, 22*(3), 363–375.

Luczak, S. E., Wall, T. L., Cook, T. A., Shea, S. H., & Carr, L. G. (2004). ALDH2 status and conduct disorder mediate the relationship between ethnicity and alcohol dependence in Chinese, Korean, and White American college students. *Journal of Abnormal Psychology, 113*(2), 271–278.

Lynam, D. R., Milich, R., Zimmerman, R., Novak, S. P., Logan, T., Martin, C.,

et al. (1999). Project DARE: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology, 67*(4), 590.

MacKillop, J., & Lisman, S. A. (2008). Effects of a context shift and multiple context extinction on reactivity to alcohol cues. *Experimental and Clinical Psychopharmacology, 16*(4), 322–331.

MacKillop, J., & Monti, P. M. (2007). Advances in the scientific study of craving for alcohol and tobacco: From scientific study to clinical practice. In P. M. Miller & D. J. Kavanagh (Eds.), *Translation of addictions sciences into practice.* Amsterdam: Elsevier.

Maisto, S. A., O'Farrell, T. J., Connors, G. J., McKay, J. R., & Pelcovits, M. (1988). Alcoholics' attributions of factors affecting their relapse to drinking and reasons for terminating relapse episodes. *Addictive Behaviors, 13*(1), 79–82.

Marlatt, A. G., & Gordon, J. R. (1985). Relapse prevention: Maintenance strategies in the treatment of addictive behaviors. New York: Guilford Press.

Marlatt, A. G., Larimer, M. E., Baer, J. S., & Quigley, L. A. (1993). Harm reduction for alcohol problems: Moving beyond the controlled drinking controversy. *Behavior Therapy, 24*(4), 461–503.

Mason, B. J., & Heyser, C. J. (2010). Acamprosate: A prototypic neuromodulator in the treatment of alcohol dependence. *CNS and Neurological Disorders Drug Targets, 9*(1), 23–32.

McCaul, M. E., Stitzer, M. L., Bigelow, G. E., & Liebson, I. A. (1984). Contingency management interventions: Effects on treatment outcome during methadone detoxification. *Journal of Applied Behavior Analysis, 17*(1), 35–43.

McCaul, M. E., Wand, G. S., Eissenberg, T., Rohde, C. A., & Cheskin, L. J. (2000). Naltrexone alters subjective and psychomotor responses to alcohol in heavy drinking subjects. *Neuropsychopharmacology, 22*(5), 480–492.

McCrady, B. S. (2000). Alcohol use disorders and the Division 12 Task Force of the American Psychological Association. *Psychology of Addictive Behaviors, 14*(3), 267.

McCrady, B. S., Epstein, E. E., Cook, S., Jensen, N., & Hildebrandt, T. (2009). A randomized trial of individual and couple behavioral alcohol treatment for women. *Journal of Consulting and Clinical Psychology, 77*(2), 243.

McCrady, B. S., Epstein, E. E., & Hirsch, L. S. (1996). Issues in the implementation of a randomized clinical trial that includes Alcoholics Anonymous: Studying AA-related behaviors during treatment. *Journal of Studies on Alcohol and Drugs, 57*(6), 604.

McCrady, B. S., Epstein, E. E., & Hirsch, L. S. (1999). Maintaining change after conjoint behavioral alcohol treatment for men: Outcomes at 6 months. *Addiction, 94*(9), 1381–1396.

McCrady, B. S., Epstein, E. E., & Kahler, C. W. (2004). Alcoholics Anonymous and relapse prevention as maintenance strategies after conjoint behavioral alcohol treatment for men: 18-month outcomes. *Journal of Consulting and Clinical Psychology, 72*(5), 870–878.

McCrady, B. S., Horvath, T. A., & Delaney, S. I. (2003). *Self-help groups* (3rd ed.). Boston: Allyn & Bacon.

McDonell, M. G., Howell, D. N., McPherson, S., Cameron, J. M., Srebnik, D., Roll, J. M., et al. (2012). Voucher-based reinforcement for alcohol abstinence using the ethyl-glucuronide alcohol biomarker. *Journal of Applied Behavior Analysis, 45*(1), 161–165.

Mello, N. K., & Mendelson, J. H. (1965). Operant analysis of drinking patterns of chronic alcoholics. *Nature, 206*, 43–46.

Merrill, J. C., Pinsky, I., Killeya-Jones, L. A., Sloboda, Z., & Dilascio, T. (2006). Substance abuse prevention infrastructure: A survey-based study of the organizational structure and function of the DARE program. *Substance Abuse Treatment, Prevention, and Policy, 1*(25), 13.

Meyers, R. J., Miller, W. R., Hill, D. E., & Tonigan, J. S. (1998). Community reinforcement and family training (CRAFT): Engaging unmotivated drug users in treatment. *Journal of Substance Abuse, 10*(3), 291–308.

Meyers, R. J., Roozen, H. G., & Smith, J. E. (2011). The community reinforcement approach: An update of the evidence. *Alcohol Research and Health, 33*(4), 380.

Meyers, R. J., & Smith, J. E. (1995). *Clinical guide to alcohol treatment: The community reinforcement approach*. New York: Guilford Press.

Meyers, R. J., & Smith, J. E. (1997). Getting off the fence: Procedures to engage treatment-resistant drinkers. *Journal of Substance Abuse Treatment, 14*(5), 467–472.

Miller, W. R. (1996). What is a relapse?: Fifty ways to leave the wagon. *Addiction, 91*(12S1), S15–S27.

Miller, W. R., & Baca, L. M. (1983). Two-year follow-up of bibliotherapy and therapist-directed controlled drinking training for problem drinkers. *Behavior Therapy, 14*(3), 441–448.

Miller, W. R., Benefield, R. G., & Tonigan, J. S. (1993). Enhancing motivation for change in problem drinking: A controlled comparison of two therapist styles. *Journal of Consulting and Clinical Psychology, 61*(3), 455–461.

Miller, W. R., & Hester, R. K. (1995). Treatment for alcohol problems: Toward an informed eclecticism. In R. K. Hester & W. R. Miller (Eds.), *Handbook of alcoholism treatment approaches: Effective alternatives* (2nd ed., pp. 1–12). Boston: Allyn & Bacon.

Miller, W. R., Leckman, A. L., Delaney, H. D., & Tinkcom, M. (1992). Long-term follow-up of behavioral self-control training. *Journal of Studies on Alcohol and Drugs, 53*(3), 249–261.

Miller, W. R., Meyers, R. J., & Tonigan, J. S. (1999). Engaging the unmotivated in treatment for alcohol problems: A comparison of three strategies for intervention through family members. *Journal of Consulting and Clinical Psychology, 67*(5), 688–697.

Miller, W. R., & Muñoz, R. F. (2004). *Controlling your drinking: Tools to make moderation work for you*. New York: Guilford Press.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. New York: Guilford Press.

Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American Psychologist, 64*(6), 527–537.

Miller, W. R., Taylor, C. A., & West, J. A. C. (1980). Focused versus broad-spectrum behavior therapy for problem drinkers. *Journal of Consulting and Clinical Psychology, 48*(5), 590.

Miller, W. R., Walters, S. T., & Bennett, M. E. (2001). How effective is alcoholism treatment in the United States? *Journal of Studies on Alcohol and Drugs, 62*(2), 211–220.

Miller, W. R., Westerberg, V. S., Harris, R. J., & Tonigan, J. S. (1996). What

predicts relapse?: Prospective testing of antecedent models. *Addiction, 91*(12S1), S155–S172.

Miller, W. R., & Wilbourne, P. L. (2002). Mesa Grande: A methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction, 97*(3), 265–277.

Miller, W. R., Wilbourne, P. L., & Hettema, J. E. (2003). What works?: A summary of alcohol treatment outcome research. *Handbook of Alcoholism Treatment Approaches, 3*, 13–63.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology, 72*(6), 1050–1062.

Monti, P. M., Abrams, D. B., Binkoff, J. A., Zwick, W. R., Liepman, M. R., Nirenberg, T. D., et al. (1990). Communication skills training, communication skills training with family and cognitive behavioral mood management training for alcoholics. *Journal of Studies on Alcohol and Drugs, 51*(3), 263–270.

Monti, P. M., Kadden, R., Rohsenow, D. J., Cooney, N. & Abrams, D. B. (2002). *Treating alcohol dependence: A coping skills training guide* (2nd ed.). New York: Guilford Press.

Monti, P. M., Rohsenow, D. J., Rubonis, A. V., Niaura, R. S., Sirota, A. D., Colby, S. M., et al. (1993). Cue exposure with coping skills treatment for male alcoholics: A preliminary investigation. *Journal of Consulting and Clinical Psychology, 61*(6), 1011–1019.

Monti, P. M., Rohsenow, D. J., Swift, R. M., Gulliver, S. B., Colby, S. M., Mueller, T. I., et al. (2001). Naltrexone and cue exposure with coping and communication skills training for alcoholics: Treatment process and 1-year outcomes. *Alcoholism: Clinical and Experimental Research, 25*(11), 1634–1647.

Moos, R. H., Finney, J. W., & Cronkite, R. C. (1990). *Alcoholism treatment: Context, process, and outcome*. New York: Oxford University Press.

Moos, R. H., & Moos, B. S. (2006). Participation in treatment and Alcoholics Anonymous: A 16-year follow-up of initially untreated individuals. *Journal of Clinical Psychology, 62*(6), 735–750.

Moyer, A., Finney, J. W., Swearingen, C. E., & Vergun, P. (2002). Brief interventions for alcohol problems: A meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction, 97*(3), 279–292.

Moyers, T. B., & Martin, T. (2006). Therapist influence on client language during motivational interviewing sessions. *Journal of Substance Abuse Treatment, 30*(3), 245–251.

Moyers, T. B., Martin, T., Christopher, P. J., Houck, J. M., Tonigan, J. S., & Amrhein, P. C. (2007). Client language as a mediator of motivational interviewing efficacy: Where is the evidence? *Alcoholism: Clinical and Experimental Research, 31*(s3), 40s–47s.

Moyers, T. B., Martin, T., Houck, J. M., Christopher, P. J., & Tonigan, J. S. (2009). From in-session behaviors to drinking outcomes: A causal chain for motivational interviewing. *Journal of Consulting and Clinical Psychology, 77*(6), 1113.

Moyers, T. B., Miller, W. R., & Hendrickson, S. M. L. (2005). How does motivational interviewing work?: Therapist interpersonal skill predicts client

involvement within motivational interviewing sessions. *Journal of Consulting and Clinical Psychology, 73*(4), 590.

Mundt, M. P., Parthasarathy, S., Chi, F. W., Sterling, S., & Campbell, C. I. (2012). 12-Step participation reduces medical use costs among adolescents with a history of alcohol and other drug treatment. *Drug and Alcohol Dependence, 126*(1–2), 124–130.

Noel, N. E., & McCrady, B. S. (1993). Alcohol-focused spouse involvement with behavioral marital therapy. In T. J. O'Farrell (Ed.), *Treating alcohol problems: Marital and family interventions* (pp. 210–235). New York: Guilford Press.

O'Farrell, T. J. (1995). *Marital and family therapy.* In R. Hester & W. R. Miller (Eds.), *Handbook of alcoholism treatment approaches* (2nd ed., pp. 195–220). Boston: Allyn & Bacon.

O'Farrell, T. J., & Birchlery, G. R. (1987). Marital relationships of alcoholic, conflicted, and nonconflicted couples. *Journal of Marital and Family Therapy, 13*(3), 259–274.

O'Farrell, T. J., Cutter, H. S. G., Choquette, K. A., Floyd, F. J., & Bayog, R. D. (1992). Behavioral marital therapy for male alcoholics: Marital and drinking adjustment during the two years after treatment. *Behavior Therapy, 23*(4), 529–549.

O'Farrell, T. J., Cutter, H. S. G., & Floyd, F. J. (1985). Evaluating behavioral marital therapy for male alcoholics: Effects on marital adjustment and communication from before to after treatment. *Behavior Therapy, 16*(2), 147–167.

O'Malley, S. S., Garbutt, J. C., Gastfriend, D. R., Dong, Q., & Kranzler, H. R. (2007). Efficacy of extended-release naltrexone in alcohol-dependent patients who are abstinent before treatment. *Journal of Clinical Psychopharmacology, 27*(5), 507–512.

Oei, T. P. S., & Jackson, P. R. (1980). Long-term effects of group and individual social skills training with alcoholics. *Addictive Behaviors, 5*(2), 129–136.

Oei, T. P. S., & Jackson, P. R. (1982). Social skills and cognitive behavioral approaches to the treatment of problem drinking. *Journal of Studies on Alcohol and Drugs, 43*(5), 532.

Orford, J., & Keddie, A. (1986). Abstinence or controlled drinking in clinical practice: A test of the dependence and persuasion hypotheses. *British Journal of Addiction, 81*(4), 495–504.

Ouimette, P. C., Moos, R. H., & Finney, J. W. (1998). Influence of outpatient treatment and 12-step group involvement on one-year substance abuse treatment outcomes. *Journal of Studies on Alcohol and Drugs, 59*(5), 513–522.

Pan, W., & Bai, H. (2009). A multivariate approach to a meta-analytic review of the effectiveness of the DARE program. *International Journal of Environmental Research and Public Health, 6*(1), 267–277.

Petry, N. M., Alessi, S. M., & Ledgerwood, D. M. (2012). A randomized trial of contingency management delivered by community therapists. *Journal of Consulting and Clinical Psychology, 80*(2), 286–298.

Petry, N. M., Martin, B., Cooney, J. L., & Kranzler, H. R. (2000). Give them prizes, and they will come: Contingency management for treatment of alcohol dependence. *Journal of Consulting and Clinical Psychology, 68*(2), 250–257.

Pettinati, H. M., O'Brien, C. P., Rabinowitz, A. R., Wortman, S. P., Oslin, D. W., Kampman, K. M., et al. (2006). The status of naltrexone in the treatment of

alcohol dependence: Specific effects on heavy drinking. *Journal of Clinical Psychopharmacology, 26*(6), 610–625.

Pettinati, H. M., Silverman, B. L., Battisti, J. J., Forman, R., Schweizer, E., & Gastfriend, D. R. (2011). Efficacy of extended-release naltrexone in patients with relatively higher severity of alcohol dependence. *Alcoholism: Clinical and Experimental Research, 35*(10), 1804–1811.

Prendergast, M., Podus, D., Finney, J., Greenwell, L., & Roll, J. (2006). Contingency management for treatment of substance use disorders: A meta-analysis. *Addiction, 101*(11), 1546–1560.

Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol and Drugs, 58*(1), 7–29.

Project MATCH Research Group. (1998). Matching alcoholism treatments to client heterogeneity: Project MATCH three-year drinking outcomes. *Alcoholism: Clinical and Experimental Research, 22*(6), 1300–1311.

Ramo, D. E., & Brown, S. A. (2008). Classes of substance abuse relapse situations: A comparison of adolescents and adults. *Psychology of Addictive Behaviors, 22*(3), 372–379.

Rogers, C. R. (1959). A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: The study of a science.* Vol. 3. *Formulations of the person and the social context* (pp. 184–256). New York: McGraw-Hill.

Rohsenow, D. J., Monti, P. M., Hutchison, K. E., Swift, R. M., Colby, S. M., & Kaplan, G. B. (2000). Naltrexone's effects on reactivity to alcohol cues among alcoholic men. *Journal of Abnormal Psychology, 109*(4), 738–742.

Rohsenow, D. J., Monti, P. M., Rubonis, A. V., Gulliver, S. B., Colby, S. M., Binkoff, J. A., et al. (2001). Cue exposure with coping skills training and communication skills training for alcohol dependence: 6- and 12-month outcomes. *Addiction, 96*(8), 1161–1174.

Roizen, R. (1987). The great controlled-drinking controversy. *Recent Developments in Alcoholism, 5*, 245–279.

Roman, P. M., & Johnson, J. A. (1998). *Private sector programs: On-site interview results* (NTCS Report No. 3, 1998). Athens: University of Georgia.

Roozen, H. G., Boulogne, J. J., van Tulder, M. W., van den Brink, W., De Jong, C. A., & Kerkhof, A. J. (2004). A systematic review of the effectiveness of the community reinforcement approach in alcohol, cocaine and opioid addiction. *Drug and Alcohol Dependence, 74*(1), 1–13.

Rosenbaum, D. P., Flewelling, R. L., Bailey, S. L., Ringwalt, C. L., & Wilkinson, D. L. (1994). Cops in the classroom: A longitudinal evaluation of Drug Abuse Resistance Education (DARE). *Journal of Research in Crime and Delinquency, 31*(1), 3–31.

Rosenbaum, D. P., & Hanson, G. S. (1998). Assessing the effects of school-based drug education: A six-year multilevel analysis of Project DARE. *Journal of Research in Crime and Delinquency, 35*(4), 381–412.

Rosenberg, H. (1993). Prediction of controlled drinking by alcoholics and problem drinkers. *Psychological Bulletin, 113*(1), 129–139.

Rosenberg, H., & Davis, L. A. (1994). Acceptance of moderate drinking by alcohol treatment services in the United States. *Journal of Studies on Alcohol and Drugs, 55*(2), 167–172.

Rösner, S., Hackl-Herrwerth, A., Leucht, S., Lehert, P., Vecchi, S., & Soyka, M. (2010). Acamprosate for alcohol dependence. *Cochrane Database of Systematic Reviews, 9*.

Rotgers, F., Kern, M. F., & Hoeltzel, R. (2002). *Responsible drinking: A moderation management approach for problem drinkers*. Oakland, CA: New Harbinger.

Ryan, R. M., Plant, R. W., & O'Malley, S. (1995). Initial motivations for alcohol treatment: Relations with patient characteristics, treatment involvement, and dropout. *Addictive Behaviors, 20*(3), 279–297.

Saladin, M. E., & Santa Ana, E. J. (2004). Controlled drinking: More than just a controversy. *Current Opinion in Psychiatry, 17*(3), 175–187.

Sanchez-Craig, M., & Lei, H. (1986). Disadvantages to imposing the goal of abstinence on problem drinkers: An empirical study. *British Journal of Addiction, 81*(4), 505–512.

Satcher, D. (2001). The Surgeon General's call to action to promote sexual health and responsible sexual behavior. *American Journal of Health Education, 32*(6), 356–368.

Schoener, E. P., Madeja, C. L., Henderson, M. J., Ondersma, S. J., & Janisse, J. J. (2006). Effects of motivational interviewing training on mental health therapist behavior. *Drug and Alcohol Dependence, 82*(3), 269–275.

Sitharthan, T., Sitharthan, G., Hough, M. J., & Kavanagh, D. J. (1997). Cue exposure in moderation drinking: A comparison with cognitive-behavior therapy. *Journal of Consulting and Clinical Psychology, 65*(5), 878–882.

Skinner, B. F. (1938/1991). *The behavior of organisms: An experimental analysis*. Acton, MA: Copley Publishing Group.

Smith, D. C., Godley, S. H., Godley, M. D., & Dennis, M. L. (2011). Adolescent community reinforcement approach outcomes differ among emerging adults and adolescents. *Journal of Substance Abuse Treatment, 41*(4), 422–430.

Smith, D. I. (1985). Evaluation of a residential AA program for women. *Alcohol and Alcoholism, 20*, 315–327.

Smith, D. I. (1986). Evaluation of a residential AA program. *Substance Use and Misuse, 21*(1), 33–49.

Smith, J. E., & Meyers, R. J. (1995). *The community reinforcement approach*. Boston: Allyn & Bacon.

Smith, J. E., Meyers, R. J., & Delaney, H. D. (1998). The community reinforcement approach with homeless alcohol-dependent individuals. *Journal of Consulting and Clinical Psychology, 66*(3), 541–548.

Sobell, L. C., Cunningham, J. A., & Sobell, M. B. (1996). Recovery from alcohol problems with and without treatment: Prevalence in two population surveys. *American Journal of Public Health, 86*(7), 966–972.

Sobell, M. B., & Sobell, L. C. (1973). Alcoholics treated by individualized behavior therapy: One year treatment outcome. *Behaviour Research and Therapy, 11*(4), 599–618.

Sobell, M. B., & Sobell, L. C. (1976). Second year treatment outcome of alcoholics treated by individualized behavior therapy: Results. *Behaviour Research and Therapy, 14*(3), 195–215.

Sobell, M. B., & Sobell, L. C. (1995). Controlled drinking after 25 years: How important was the great debate? *Addiction, 90*(9), 1149–1153.

Stasiewicz, P. R., Brandon, T. H., & Bradizza, C. M. (2007). Effects of extinction context and retrieval cues on renewal of alcohol-cue reactivity among alcohol-dependent outpatients. *Psychology of Addictive Behaviors, 21*(2), 244–248.

Steinglass, P. (1981). The alcoholic family at home: Patterns of interaction in dry, wet, and transitional stages of alcoholism. *Archives of General Psychiatry, 38*(5), 578–584.

Stitzer, M., & Petry, N. (2006). Contingency management for treatment of substance abuse. *Annual Review of Clinical Psychology, 2*, 411–434.

Suh, J. J., Pettinati, H. M., Kampman, K. M., & O'Brien, C. P. (2006). The status of disulfiram: A half of a century later. *Journal of Clinical Psychopharmacology, 26*(3), 290–302.

Swift, R. M., Whelihan, W., Kuznetsov, O., & Buongiorno, G. (1994). Naltrexone-induced alterations in human ethanol intoxication. *American Journal of Psychiatry, 151*(10), 1463–1467.

Timko, C., Moos, R. H., Finney, J. W., & Lesar, M. D. (2000). Long-term outcomes of alcohol use disorders: Comparing untreated individuals with those in Alcoholics Anonymous and formal treatment. *Journal of Studies on Alcohol and Drugs, 61*(4), 529–540.

Tonigan, J. S., Miller, W. R., & Connors, G. J. (2000). Project MATCH Client impressions about Alcoholics Anonymous: Measurement issues and relationship to treatment outcome. *Alcoholism Treatment Quarterly, 18*, 25–41.

Tonigan, J. S., Miller, W. R., & Schermer, C. (2002). Atheists, agnostics and Alcoholics Anonymous. *Journal of Studies on Alcohol and Drugs, 63*(5), 534–541.

Tournier, R. E. (1979). Alcoholics Anonymous as treatment and as ideology. *Journal of Studies on Alcohol and Drugs, 40*(3), 230–239.

Twerski, A. J. (1983). Early intervention in alcoholism: Confrontational techniques. *Hosp Community Psychiatry, 34*(11), 1027–1030.

Vaillant, G. E. (1996). A long-term follow-up of male alcohol abuse. *Archives of General Psychiatry, 53*(3), 243–249.

Volpicelli, J. R., Rhines, K. C., Rhines, J. S., Volpicelli, L. A., Alterman, A. I., & O'Brien, C. P. (1997). Naltrexone and alcohol dependence: Role of subject compliance. *Archives of General Psychiatry, 54*(8), 737–742.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive-based approaches for weight loss: A randomized trial. *Journal of the American Medical Association, 300*(22), 2631–2637.

Volpp, K. G., Loewenstein, G., Troxel, A. B., Doshi, J., Price, M., Laskin, M., et al. (2008). A test of financial incentives to improve warfarin adherence. *BMC Health Services Research, 8*(1), 272.

Walitzer, K. S., & Dermen, K. H. (2004). Alcohol-focused spouse involvement and behavioral couples therapy: Evaluation of enhancements to drinking reduction treatment for male problem drinkers. *Journal of Consulting and Clinical Psychology, 72*(6), 944–955.

Walsh, D. C., Hingson, R. W., Merrigan, D. M., Levenson, S. M., Cupples, L. A., Heeren, T., et al. (1991). A randomized trial of treatment options for alcohol-abusing workers. *New England Journal of Medicine, 325*(11), 775–782.

Walters, G. D. (2001). Behavioral self-control training for problem drinkers: A meta-analysis of randomized control studies. *Behavior Therapy, 31*(1), 135–149.

White, W. L. (1998). *Slaying the dragon: A history of addiction treatment and recovery in America*. Bloomington, IL: Chestnut Health Systems/Lighthouse Institute.

White, W. L., & Miller, W. R. (2007). The use of confrontation in addiction treatment: History, science, and time for change. *Counselor, 8*, 12–30.

Wilk, A. I., Jensen, N. M., & Havighurst, T. C. (1997). Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. *Journal of General Internal Medicine, 12*(5), 274–283.

CHAPTER TWELVE

Herbal Treatments and Antidepressant Medication

Similar Data, Divergent Conclusions

Harald Walach and Irving Kirsch

Depression is one of the most widespread psychological disorders. According to statistics reported by the Centers for Disease Control and Prevention (2011), approximately 8% of adults and adolescents reported being depressed during any 2-week period between 2006 and 2008. The most common treatment for depression is antidepressant medication. From 2005 to 2008, about 1 in 10 Americans aged 12 and over took antidepressants, an increase of 400% over estimates made a decade earlier (Pratt, Brody, & Gu, 2011). During this same period, a controversy developed regarding the efficacy of these treatments. Although a strong positive response to antidepressive medication is undeniable, clinical trials indicate that most of this response is duplicated by inert placebos. This finding has led some authors to question whether antidepressants are much more than active placebos (Kirsch, 2010; Moncreiff & Kirsch, 2005).

Paralleling the development of new pharmacological treatments for depression, there has been an increased interest in some of the oldest treatments for physical and psychological disorders. As part of a more general interest in alternative and complementary medicines, the use of herbal remedies has attracted the attention of consumers, researchers, and health professionals.

In this chapter, we review the data on antidepressant medications and herbal remedies for psychological disorders. Specifically, we address the issue of the degree to which the effects of these treatments are due to their physical properties as opposed to placebo effects.

ANTIDEPRESSANT MEDICATION

Hopelessness is at the core of most cases of depression. Hopelessness is also an expectation. It is an expectancy that an intolerable situation will not improve. This being the case, one would expect depression to be very reactive to the placebo effect. Placebos instill an expectancy for improvement, and expectancy addresses a core issue of depression.

Placebos are typically used to control for the psychological effects of administering a treatment. To that end, articles reporting clinical trials and reviews of the literature typically focus on the difference between the response to placebo and the response to active treatment. The question that is conventionally addressed is whether this difference is statistically significant.The magnitude of the difference is less frequently considered, and scant attention is typically given to the magnitude of the response to placebo.

In contrast to conventional reviews, Kirsch and Sapirstein (1998, 1999) reported a meta-analysis of antidepressant medication, in which both the drug effect and the placebo effect were evaluated. Computer searches of Medline and PsychLit, supplemented by studies identified in prior reviews, produced 19 published, placebo-controlled, randomized clinical trials of the acute effects of antidepressants on patients with a clear primary diagnosis of depression. Taken together, these studies contained 2,318 participants, of whom 1,460 received medication and 858 received placebo.

Drug effects are assessed by comparing clinical outcomes in patients given medications with outcomes in patients given placebo (see also Gaudiano, Dalrymple, Weinstock, & Lohr, Chapter 6, this volume). The drug effect is assumed to be the difference between the response to the drug and the response to the placebo. Just as one has to control for the placebo effect to evaluate the drug effect, one has to control for natural history effects (e.g., spontaneous remission and regression to the mean) to evaluate placebo effects. To that end, Kirsch and Sapirstein (1998, 1999) conducted a second search to identify studies in which changes in depression were reported for patients assigned to no-treatment or wait-list control groups. This produced an additional 19 studies containing 244 patients who had been assigned to wait-list or no-treatment control groups. Changes in depression among these patients were used as a baseline, against which the response to placebo could be compared. Analyses of patient characteristics indicated that participants in the two groups of studies were comparable in terms of age, duration of treatment (or length of time between assessments for the no-treatment groups), and pretreatment scores on the Hamilton Rating Scale for depression and the Beck Depression Inventory.

The effect size (*d*; see Glossary) for pretreatment to posttreatment changes in depression was evaluated for each of these three groups of patients (those randomized to active antidepressant, those randomized to placebo, and those randomized to no-treatment or wait-list control). In

studies reporting multiple measures of depression, an effect size was calculated for each measure, and these effect sizes were then averaged. In studies reporting the effects of two drugs, a single mean effect size for both was calculated. Thus, there were 19 effect sizes for each condition (drug, placebo, and no-treatment).

The pre–post effect size for antidepressant drugs was 1.55 standard deviations. This is a very large effect size (Cohen, 1988), and it indicates that administration of an antidepressant medication is followed by substantial clinical improvement. However, the pre–post effect size for response to placebo was 1.16. This finding indicates that 75% of the effect of antidepressant medication can be duplicated by administration of an inert placebo. In contrast, analysis of the course of untreated depression over the same time period indicated an effect size of only 0.37 standard deviations. Taken together, these effect sizes suggest that about 25% of the response to antidepressant medication may be a true drug effect, another 25% may be due to the natural history of the condition, and about 50% is an expectancy effect.

Despite the magnitude of the placebo effect, the data in the Kirsch and Sapirstein (1998, 1999) meta-analyses indicate a reasonably sizable advantage for the active drug over placebo. However, there is reason to believe that much of this difference may be due to expectancy rather than to the pharmacological properties of the drugs. Kirsch and Sapirstein subdivided the set of studies by type of medication (e.g., tricyclics, selective serotonin reuptake inhibitors [SSRIs], monoamineoxidase inhibitors [MAOIs]). They found that the pretreatment to posttreatment effect size was fairly consistent across drug type. More remarkable was the finding that the proportion of the effect size duplicated by placebo was virtually identical across medication type (range = 74 to 76%). The biggest surprise, however, came when Kirsch and Sapirstein examined the effect size for a subset of studies in which the active drugs (amylobarbitone, lithium, liothyronine, and adinazolam) were not standard antidepressants for acute use, but sedatives, tranquillizers, or preventive medications. The pre–post effect size for these drugs (d = 1.69) was as great as that for antidepressants, and again an inactive placebo duplicated 76% of this response.

It is possible that amylobarbitone, lithium, liothyronine, and adinazolam are in fact antidepressants, with pharmacological effects as great as tricyclics, SSRIs, MAOIs, and the others. Alternatively, it is possible that all of these drugs function as active placebos. An active placebo is an active medication that does not have specific activity for the condition being treated, but that mimics the side effects of the active drug that is being investigated. Fisher and Greenberg (1989) summarized data indicating that the effect of antidepressant medication is smaller when it is compared with an active placebo than when compared with an inert placebo. Quitkin, Rabkin, Gerald, Davis, and Klein (2000) reanalyzed those data and came to an opposite conclusion. In fact, relatively few studies assess

this possibility, and no studies assess it in relation to the more recent and most recently prescribed antidepressants (e.g., the SSRIs).

The active placebo hypothesis is based on the idea that patients randomized to the active arm of clinical trials are able to deduce that they have been assigned to the drug condition because the active drugs produce more side effects than inert placebos. The ability of patients and physicians to detect assignment to the drug condition has been well established, not only for antidepressants (Blashki, Mowbray, & Davies, 1971; Rabkin et al., 1986), but also other medications (Baethge, Assall, & Baldessarini, 2013; Quitkin et al., 2000). The penetration of the double blind may produce an enhanced placebo effect in drug conditions and a diminished placebo effect in placebo groups. Thus, the apparent drug effect of antidepressants may in part be a placebo effect, magnified by differences in experienced side effects and patients' subsequent recognition of the condition to which they have been assigned. Support for this interpretation is provided by a meta-analysis showing substantial correlations between unblinding and drug–placebo differences in clinical trials of psychotropic medication (Baethge et al., 2013).

Given these data, it is not surprising that the Kirsch and Sapirstein meta-analysis became the focus of considerable controversy. Critics raised concerns about (1) the relatively small number of studies evaluated, considering the large body of literature evaluating antidepressant medication and (2) various aspects of the statistical analyses (e.g., Dawes, 1998; Klein, 1998). However, other meta-analyses (Gerson, Belin, Kaufman, Mintz, & Jarvik, 1999; Joffe, Sokolov, & Streiner, 1996; Kahn, Warner, & Brown, 2000; Walach & Maidhof, 1999; Walach, Sadaghiani, Dehm, & Bierman, 2005), conducted by different authors on different sets of studies and using different statistical procedures, revealed pre–post drug and placebo effect sizes very similar to those reported by Kirsch and Sapirstein (1998, 1999). The close correspondence in the results of these independently conducted meta-analyses, despite little or no overlap in the studies included for analysis (only two studies were included in both the Joffe et al. meta-analysis and the Kirsch and Sapirstein meta-analysis), suggest that their findings and conclusions are robust.

Given the controversy stimulated by the Kirsch and Sapirstein (1998, 1999) meta-analysis, Kirsch and colleagues sought to replicate it using a different dataset (Kirsch et al., 2002, 2008). To that end, they used the Freedom of Information Act to request that the Food and Drug Administration (FDA) send them the data that pharmaceutical companies had sent to it in the process of obtaining approval for six new-generation antidepressants that accounted for the bulk of antidepressant prescriptions being written at the time. The FDA dataset has a number of advantages. First, the FDA requires that the pharmaceutical companies provide information on all the clinical trials they have sponsored. Thus, Kirsch et al. (2002, 2008) had data on unpublished trials as well as published trials. Second, the same

primary outcome measure—the Hamilton Depression Scale (HAM-D)—was used in all of the trials. That made it easy to understand the clinical significance of the drug–placebo differences. Third, these were the data on the basis of which the medications were approved. In that sense, they have a privileged status. If there is anything wrong with them, the decision to approve the medications in the first place can be called into question.

In the data sent by the FDA, only 43% of the trials showed a statistically significant benefit of drug over placebo. The remaining 57% were failed or negative trials. The results of the Kirsch et al. (2002, 2008) analyses indicated that the placebo response was 82% of the response to these antidepressants. More importantly, the mean difference between drug and placebo was less than two points on the HAM-D, and the standardized mean difference (*SMD*) was 0.32. The National Institute for Clinical Excellence (NICE), which drafts treatment guidelines for the National Health Service in the United Kingdom, has established a 3-point difference between drug and placebo on the HAM-D or an *SMD* of 0.50 as the criteria of clinical significance (NICE, 2004). Considering that the HAM-D is a scale ranging from 0 to 51 points, on which a 6-point difference can be obtained merely by changes in sleep pattern, a 3-point difference seems a rather lenient criterion. Thus, when published and unpublished data are combined, they fail to show a clinically significant advantage for antidepressant medication over inert placebo. Our analyses have since been replicated repeatedly, and despite differences in the way the data are spun, they numbers are remarkably consistent. *SMD*s range between 0.30 and 0.34, and the raw score differences on the HAM-D are always below 3 points (Fournier et al., 2010; NICE, 2004; Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008).

The meta-analyses already described were limited to studies of the acute effects of antidepressant drugs and placebos (the mean duration of the studies was 5 weeks). Walach and Maidhof (1999) extended these findings to long-term effects (6 months to 3 years). In the most stringent analysis of their data (reported in Kirsch, 1998), confined to studies in which dropouts were analyzed as treatment failures, the results indicated that 73% of the long-term improvement among patients treated with antidepressants was duplicated in patients treated by placebo. Similarly, the meta-analyses conducted by NICE (2004) as part of the process of drafting clinical guidelines for the treatment of depression also found that the results of long-term trials were similar to those of short-term trials. Specifically, they reported an *SMD* of 0.34 for the short-term trials and 0.28 for the long-term trials, both effect sizes being well short of that needed for clinical significance.

The participants in clinical trials differ from those in clinical practice in a number of ways. Typically, those who are chronic, comorbid, treatment-resistant, or responsive to placebo are screened out. These exclusion criteria render almost 80% of depressed patients ineligible for participation in standard clinical trials of antidepressants (Wisniewski et al., 2009). To

evaluate the generalizability of clinical trial data to clinical practice, the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial used broad inclusion and minimal exclusion criteria (Rush et al., 2006). Their data showed that response and remission rates were substantially higher—50% and 58%, respectively—in patients included in clinical trials than in patients who would have been excluded (Wisniewski et al., 2009). Thus, clinical trial data show only modest benefits of antidepressants over placebo, and even these small benefits are greater than those seen in clinical practice. Despite data showing the contrary, the received wisdom of the health industry is that antidepressant medication is exceptionally effective and should be the first line of attack in the treatment of depression. In the second part of this chapter we turn to another set of somatic treatments—herbal remedies—for which similar data have led to very different conclusions.

HERBAL TREATMENTS FOR PSYCHOLOGICAL DISORDERS

Among unconventional or complementary and alternative medicines, herbal treatments (phytotherapeutics; see Glossary) are increasingly popular (Chrubasik, Junck, Zappe, & Stutzke, 1998; LaFrance et al., 2000; Pirotta, Cohen, Kotsirilos, & Farish, 2000; Sparber et al., 2000; Eardley et al., 2012). This popularity can be seen as part of a public trend toward natural, holistic, or ecological approaches to life in general and to health care in particular (Furnham & Kirkcaldy, 1996; Nissen, Schunder-Tatzber, Weidenhammer, & Johannessen, 2012). By this we mean orientations in life that opt for the integration of mind and body, and that argue that scientific approaches should not only cover analyses of wholes into constituent parts but also the knowledge of how those constituents cooperate to produce a synergistic effect (Pincus & Walach, 2012). This orientation seems to reflect a subtle shift in the lay perceptions about life and values in general, or a kind of "zeitgeist." It is the expression of disillusion with the technological and scientific approach to life in general and health in particular. Many people have become afraid that the high-tech medical approach of the Western countries might lose the human aspects of care and the whole person, and focus only on diseased organs and mechanisms. This ecological stance or world-view expressing appreciation for nature and its intricate interconnectedness can become an unreflected mode of thinking and a mindset that has aptly been called the "natural commonplace" (Pratkanis, 1995), which reflects the bias that everything that is natural is good, that nature is better than culture, and that technology per se is antithetical to life.

The fitting together of belief systems in patients and therapists is part and parcel of maximizing nonspecific treatment factors (Frank, 1987). Thus, people subscribing to an ecological and holistic worldview are likely

to seek out treatments consistent with those views. For instance, people who believe that the universe is permeated by some invisible life-force are likely to be attracted to acupuncture or traditional Chinese medicine with its teaching that life is upheld by the flow and balance of "qi." Similarly, someone who is convinced that well-being is dependent on an intimate functioning and interconnection of mental and bodily systems will probably be attracted to the system of homeopathy, which takes into account not only physical symptoms of the present disease, but also psychological symptoms and the personality as a whole. Data indicate that patients who seek out complementary and alternative treatments are not dissatisfied with the outcomes of conventional medical treatments. Instead, it is primarily the ideological underpinnings of complementary and alternative treatments that seem attractive to these patients (Astin, 1998). In addition, the side effects that are common in effective conventional treatments are often cited as reasons for seeking out alternatives (Walach & Güthlin, 2000).

Hence the turn toward naturopathic treatments, and among them phytotherapeutic medicines (i.e., made from herbal substances; see Glossary), is a logical consequence of the turn toward nature in the general zeitgeist. The same is true not only for patients but for doctors: Doctors who are younger, have been treated by alternative and complementary medical approaches themselves, and adhere to a more holistic view of medicine are likely to attribute more power to naturopathic interventions (Easthope, Tranter, & Gill, 2001).

Phytotherapeutic approaches are often seen as harmless and imbued with fewer side effects than conventional treatments, which probably is an extension of the "natural commonplace," which we defined earlier (Pratkanis, 1995). In scientific circles, the inverse of this natural commonplace is frequently popular: Everything that is "natural" and not "scientific" is likely to have small effects. Both assumptions may be mistaken in two respects. First, because phytotherapeutic interventions fit some patients' worldviews better, they may be more likely to trigger positive expectancy effects. Often labeled as placebo effects, these expectancy effects are frequently dismissed and ignored, but they can be of substantial benefit. As discussed earlier, they have been reported to share between 60 and 80% of the variance of pharmacological treatment effects in depression (Kirsch & Sapirstein, 1998, 1999; Walach & Maidhof, 1999) and other disorders (Walach et al., 2005), thus potentially contributing considerably to the efficacy of pharmacological treatment. Second, as our review will indicate, placebo-controlled trials have shown that at least some of the more widely used phytotherapeutic methods exert specific effects over and above those attributable to placebo.

Phytotherapeutics have been widely used in Germany for a long time, where they have a specific legal status. The pharmaceuticals derived from plants or with contents derived from plants cover about 20–30% of the market in Germany, which means $2 billion (Marstedt & Moebus, 2002).

In the United States, herbal treatments are second among alternative therapies (after prayer), used by 17.7% of the population, with a seven-fold rise in use since 1990 (Barnes, Bloom, & Nahin, 2008). A conservative estimate of global sales of natural products was $31 billion in 2002 (Cordell, 2002).

The preparation with the highest sales in both prescriptions and over the counter in Germany is a phytotherapeutic remedy made from *Ginkgo biloba*, which is discussed later in this chapter (Rosslenbroich & Saller, 1992). Phytotherapeutics have made their way into drugstores and pharmacies around the world and are now major parts of the rising business of complementary and alternative medicine (Mills, 2001).

Hypericum

Preparations made of St. John's wort (hypericum) were used by traditional German naturopaths as treatments for depression. These days, hypericum is one of the most widely used phytotherapeutic compounds in Germany for the treatment of depression (Kasper, Gastpar, Möller, et al., 2010). It has been used in folk medicine since the Middle Ages to treat bruises (as a bandage with the flowers preserved in oil or fat) or as a decoction in the form of tea to treat depression. Modern pharmacology has determined that the effective components, among them the flavonoids hypericin and hyperforin, show a variety of effects comparable with those of SSRIs (Greeson, Sanford, & Monti, 2001; Singer, Wonnemann, & Müller, 1999; Access Economics, 2010; Kasper, Gastpar, Möller, et al., 2010). In contrast to artificially manufactured SSRIs, hypericin does not inhibit reuptake of serotonin (5-HT). Instead, it may induce an expression of 5-HT receptors centrally and peripherally. The latter effects would, incidentally, explain the use of hypericum in folk medicine for cuts, bruises, and lacerations. An enhanced peripheral action of 5-HT would certainly explain its pain-killing and perhaps also immunological effects. Peripheral 5-HT receptors and/or enhanced availability of serotonin play an important role in the mediation of inflammation and pain perception. 5-HT receptors can be expressed by nearly every immunocompetent cell. In this way, serotonin mediates immune responses. In contrast, cytokines, which are the substances used by the immune system to communicate with itself and with other parts of the organisms and which are produced when infections occur, induce higher bioavailability of serotonin in some parts of the brain (Mössner & Lesch, 1998).

One problem common to all phytotherapy has been researched extensively with hypericum preparations: Depending on the time of harvest, the origin of plants, and the method of extraction, the active compounds vary greatly in quality and quantity. This finding may explain, at least in part, the divergent results of clinical trials using different dosages (Laakmann, Schüle, Baghai, & Kieser, 1998), especially considering that hypericum preparations are standardized in some countries such as Germany, where

it is also regulated as a pharmaceutical and may be dispensed in different nonstandardized dosages in countries where it is regulated as a food supplement.

One of the major tasks of phytopharmacology is to identify and analyze active ingredients and to determine the best way of extracting them. Until recently, this had been done through clinical experience rather than systematic empirical study. In the case of hypericum, however, there have been considerable recent advances. A series of meta-analyses have confirmed the likelihood of a specific antidepressive action of St. John's wort (Field, Monti, Greeson, & Kunkel, 2000; Linde, Berner, Egger, & Mulrow, 2005; Linde, Berner, & Kriston, 2008).

One of the first meta-analyses concluded that hypericum extracts are effective for mild to moderate depression and that their therapeutic effects are superior to those of placebo and comparable with those of tricyclics, but with fewer side effects (Linde et al., 1996). This study included 23 randomized trials in 1,757 outpatients with mild to moderate depression. Fifteen of those trials compared hypericum to a placebo, and eight with active medication. The main outcome measure of the meta-analysis was a pooled responder rate ratio (RR; this is an effect-size measure expressed as a ratio of responders in the treated by responders in the control group). The analysis yielded a significant RR of 2.67 compared with placebo and a RR of 1.10 with a single preparation or 1.52 with combination preparations of hypericum compared with conventional medications. In a recent update of their original review, more recently published as a Cochrane review following the strict criteria of the Cochrane collaboration, Linde, Berner, and Kriston (2008) analyzed 29 trials, including 18 comparing hypericum with placebo and 17 comparisons with SSRIs or tri- or tetracyclics. The reviewers reported a significant rate ratio of 1.28 (95% confidence intervals [CIs] 1.10 to 1.49) compared with placebo and equality with tri/tetracyclics and SSRIs (RR = 1.02 or RR = 1.00, respectively; CIs 0.90 to 1.15; 0.90 to 1.12). Significantly more patients dropped out of the study when randomized to SRRIs or tri/tetracyclics compared with hypericum due to side effects (odds ratio [OR] = 0.24; CIs 0.13 to 0.46 for tri/tetracyclics and OR = 0.53; CIs 0.34 to 0.83 for SSRIs). Interestingly, the effects were stronger and more homogeneous for studies from German-speaking countries. The latter finding may have to do with the fact either that psychiatric care is different across health care systems or that hypericum as a treatment is much more part of the culture and thus produces better effects in German-speaking countries. Thus, the authors concluded that hypericum (1) is more effective than placebo, (2) is as effective as standard pharmaceutical interventions, and (3) produces fewer side effects, a finding that is backed by the recent Australian Health Technology Report (Access Economics, 2010). Another meta-analysis, based on fewer trials and comparing hypericum with SSRIs, only found equality of both treatments and significantly fewer withdrawals from hypericum arms of studies due to side effects (Rahimi, Nikfar,

& Abollahi, 2009). Also, an economic evaluation, using current data, has concluded that hypericum is more cost-effective than SSRIs, with savings between $200 and $360 per person and 0.08 to 0.12 Quality Adjusted Life Years (QALY) gained over 72 months (Solomon, Adams, & Graves, 2013) This finding supports an earlier Australian analysis (Access Economics, 2010)

One large study (Woelk et al., 2000) randomized 324 patients to receive either hypericum or imipramine for 6 weeks. Patients in both groups improved significantly, with no major between-group differences in therapeutic benefit. However, side effects were significantly more common among patients given imipramine. Side effects were reported by 63% of the patients in the imipramine group, compared with 39% of the patients in the hypericum group. Philipp, Kohnen, and Hiller (1999) reported similar results.Hypericum produced a mean improvement of 15.4 points on the Hamilton Depression Scale (HAM-D), compared with 14.2 points for imipramine and 12.1 points for placebo. In contrast, the incidence of side effects was only 0.5 events per patient for hypericum, which was not significantly different from that reported in the placebo group (0.6 events), but significantly less than that reported following imiprimine (1.2 events). Hypericum preparations also appear to have fewer side effects than SSRIs (Schulz, 2000; Linde, Berner, & Kriston, 2008; Kasper et al., 2010)

Although there are fewer side effects with hypericum extracts than with conventional antidepressants, some serious side effects have been observed. Among these are photosensitization, which has led some researchers to caution against the simple equation of "natural equals harmless." Indeed, in 2000, the German regulatory agency posted a caveat against interactions with other pharmaceuticals (Bundesinstitut für Arzneimittel und Medizinprodukte, 2000). Most side effects, however, have been observed with dosages considerably higher than standard doses and rarely are of the same severity as those of conventional preparations.Moreover, adverse effects are much more frequently seen in comparison trials with SSRIs (30–50%) than in placebo-controlled trials (3%; Greeson, Sanford, & Monti, 2001). This difference may be due, at least in part, to a greater expectancy for side effects when a conventional medication is included in the study, but it could also be due to dosage differences in the respective trials.

In a study accorded substantial coverage in the media, Shelton et al. (2001) reported a failure to find a significant difference between hypericum and placebo in 200 severely depressed patients. The response to hypericum was seen in 27% of the patients who responded to the drug and in 19% of patients who responded to placebo. Thus the treatment response is 42% greater than the response to placebo, which is comparable to that reported in meta-analyses of conventional antidepressants (e.g., Kirsch & Sapirstein, 1998, 1999). However, significantly more patients experienced complete remission with hypericum (14.3%) than with placebo (4.9%; $p = .02$). Side effects occurred in 10% or more of the patients in both groups; the most

prominent side effects were headaches and abdominal discomfort. There was a significant difference in incidence of headaches between groups: 41% of hypericum patients reported headaches compared with 25% of the placebo group ($p = .02$). Based on these data, it can be concluded either that the number of patients needed to treat to achieve one remission with hypericum in severe depression is 11 or that a sample size of 100 per group is needed to show a significant difference. When considering this result, two issues should be borne in mind (Berner, 2001): (1) Unlike other studies, in this study rather severely depressed patients were treated. The cutoff point on the Hamilton Scale was 20 (instead of 18, which is conventionally used to define mild to moderate depression), and (2) the percentage of chronically relapsing patients was 64% and thus rather high. The lower incidence of side effects may make hypericum especially useful for some patients, as it is not uncommon for 30–40% of participants to discontinue their medication during clinical trials of conventional medication (Kirsch et al., 2002).

Following the reasoning that the experience of side effects creates the expectancy of being treated with active components, we can state that the occurrence of fewer side effects should diminish the effectiveness of a drug. The lower incidence of side effects also makes the superiority of hypericum to placebo more convincing than the superiority of conventional antidepressants to placebo. Participants in clinical trials of hypericum may be less likely to perceive that they have been randomized to the active drug condition. Thus, given the comparability of hypericum side effects and placebo side effects, it is more likely that drug–placebo differences in hypericum trials reflect a genuine pharmacological effect.

Although the bulk of the data indicate a significant benefit for hypericum, at least for people with mild to moderate levels of depression, Rapaport and colleagues could not support this (Rapaport et al., 2011). Indeed, the jury regarding the efficacy and usefulness of hypericum is still out. Various large trials have been unsuccessful, and a recent study comparing hypericum to citalopram and placebo in mild depression showed no differential effect whatsoever, probably due to a large placebo effect (Rapaport et al., 2011). This negative finding may also have stemmed in part from the fact that 28% of patients in this study had some comorbidity with nondepressive conditions. Hypericum appears to be effective mainly for depressive disorders (Sarris & Kavanagh, 2009). In conclusion, optimism regarding the value of hypericum positive is tempered by negative findings, which imply that additional studies supporting the value of hypericum are required before strong claims and widespread marketing regarding its usefulness are warranted.

Ginkgo Biloba

Ginkgo biloba, a preparation made of the fresh leaves of the ginkgo tree, is another traditional phytotherapeutic drug. It is said to be invigorating,

to slow down the aging process, and to diminish the cognitive deficits of old age (i.e., memory loss and dementia symptoms). The ginkgo tree is a "fossil" among plants because it does not have a living relative among other plants and is said to be a special species of tree, midway between leaf-bearing trees and needle trees. The ginkgolides, which belong to the family of terpenoids and are part of the extracts, are unique in the plant kindgom and are contained only in ginkgo leaves. Ginkgo has been in use for these indications in Asian cultures and has conquered the Western market with what is probably the single most widely sold phytotherapeutic over-the-counter preparation in the world (Chan, Xia, & Fu, 2007; Kleijnen & Knipschild, 1992b, Rosslenbroich & Saller, 1992).

Ginkgo's widespread use stands in sharp contrast to the comparative scarcity of solid scientific data concerning its efficacy. In the first reviews of the efficacy of ginkgo for intermittent claudication (a reduction of blood flow in the peripheral and central blood vessels) and cerebral insufficiency (Kleijnen & Knipschild, 1992a, 1992b), systematic searches produced 15 controlled trials of intermittent claudication and 38 trials of cerebral insufficiency. Two of the intermittent claudication trials were reasonably well performed (i.e., randomized, double-blind, and of otherwise good methodological quality), as were eight of the cerebral insufficiency studies, and these studies were all significantly positive in favor of ginkgo. However, publication bias could not be excluded, and the authors concluded that although there was preliminary evidence for the efficacy of ginkgo in cerebral insufficiency and intermittent claudication, more and larger trials were necessary. A recent Cochrane review concluded that although there was a small effect size in favor of ginkgo, derived from 14 trials of ginkgo on walking distance, this translated only to a 64.5-meter difference on a treadmill and hence was not clinically significant. The authors also doubted the relevance of those findings because publication bias might play a role (Nicolai et al., 2013).

Fugh-Berman and Cott (1999) and Ott and Owens (1998) reviewed trials of ginkgo in dementia. These trials were placebo controlled, randomized, and conducted in patients with well-established diagnoses of either Alzheimer's disease or other forms of memory impairment, and provided some evidence of the efficacy of ginkgo for dementia. An earlier rigorous meta-analysis (Oken, Storzbach, & Kaye, 1998) found more than 50 papers, 4 of which fulfilled all inclusion criteria, among them having a clearly stated diagnosis of Alzheimer's disease, a placebo-controlled randomized design, a standardized extract of ginkgo, and objective outcome measures. Most studies failed to meet these criteria because diagnoses were insufficiently established. Three of the four studies showed significant effects. The pooled standardized mean difference was $d = 0.4$, which is a small to medium effect (Cohen, 1988) and significantly different from zero ($-p < .0001$). Based on this result, the authors concluded that ginkgo has a clear but modest effect on Alzheimer's disease. Side effects were scarce;

two cases of bleeding difficulties were reported. This analysis found some support from a recent analysis of a 20-year French longitudinal cohort, the so-called PAQUID study (Amieva et al., 2013), which analyzed 589 people who took ginkgo continuously, comparing them with a group of 149 patients on piracetam and 2,874 persons not taking anything. This epidemiological study found that people taking ginkgo prophylactically had less cognitive decline as measured with the Mini Mental State Interview.

This finding from a natural setting of one of the longest longitudinal cohorts is at odds with recent meta-analyses and the largest randomized trial to date, the GEM trial. For example, a recent Cochrane review located 36 trials that studied ginkgo preparations in patients with dementia or cognitive impairment. Although earlier trials often showed positive effects, when publication bias cannot be excluded more recent trials have been more inconsistent. Specifically, positive effects are juxtaposed against no effects in recent larger trials, prompting the authors to conclude that the usefulness of ginkgo preparations remains doubtful (Birks & Grimley Evans, 2009). This conclusion was backed by a recent update (Roland & Nergard, 2012). These results are at variance with earlier, more extensive analyses, and also with data that examined ginkgo either in natural settings or with more sensitive outcome indicators.

Recently, De Kosky and colleagues (DeKosky et al., 2008; Snitz et al., 2009) randomized 3,072 patients to ginkgo or placebo and found no meaningful difference: The rate of dementia after 7 years was 3.3 in the ginkgo group and 2.9 in the placebo group, giving a nonsignificant hazard ratio of 1.12. The analysis of this GEM trial for cognitive decline (Snitz et al., 2009) was similarly nonsignificant. The cognitive decline in both groups was comparatively small, pointing to a potential placebo effect. Such an effect, however, cannot be ascertained in the absence of any comparison standard. Still, this trial can be criticized in terms of real-world practice, as it included patients who at 79 years were comparatively old when treatment was started, 68% were obese—an important risk factor for dementia (Loef & Walach, 2012)—and 14.8% already had mild cognitive impairment. Although the GEM trial can rule out having missed even very small differences of $d = 0.01$ in its sample and over the observation period of 7 years, some still argue that those factors threaten the external and ecological validity of the data (Ihl, 2012). Also, the outcome measures—the Mini Mental State Scale and general scales of cognitive functioning, though widely used, have been criticized. When more meaningful and more sensitive outcome parameters were used, such as newly developed memory tests that assess real-world activities such as memorizing appointments or shopping lists, ginkgo was quite effective (Kaschel, 2009, 2011). Contrasting findings from short-term trials and from the real-world analysis of the PAQUID cohort suggest that ginkgo might in fact play a role in the prevention of dementia (Ihl, 2012), but this effect is difficult to pinpoint. In addition, gingko is without effect in healthy individuals as a cognitive

enhancer in normal samples, according to a meta-analysis that summarized data from 13 trials (Laws, Sweetnam, & Kondel, 2012). In summary, the consilience of evidence does not provide strong or unequivocal support for the use of gingko to minimize or prevent cognitive decline or dementia, especially in studies based on randomized controlled trials. Positive findings derived from studies in more naturalistic settings suggest that research that examines variables that potentially distinguish positive versus negative outcomes would be informative.

A review of ginkgo in coronary artery disease (which is often associated with mood disorders and other psychiatric conditions) identified eight randomized, placebo-controlled trials. Four of these trials showed a significant difference in the increase of pain-free walking distance (in a brisk walking test normally carried out over a distance of 300 feet [30 meters] until pain occurs) in favor of ginkgo, with a weighted mean difference of 36.6 meters (Gundling & Ernst, 1999). However, the GEM study analysis of cardiovascular events did not support these earlier positive findings; thus, the suggestion is that there is no difference from placebo (Kuller, Ives, Fitzpatrick, et al., 2010).

A systematic review of trials of ginkgo in patients with tinnitus identified five randomized studies, four of which were placebo-controlled (Ernst & Stevenson, 1999). Only one of those trials reported negative results, which might have been due to suboptimal doses. The other trials showed the superior effects of ginkgo to placebo. The authors of the review concluded that ginkgo seems to be an effective treatment for tinnitus, but that more and larger-scale trials are needed. The common denominator regarding gingko's mechanisms of action seems to be that it increases capillary blood flow in the brain and other organs. Kellermann and Kloft (2011), in their systematic review according to Cochrane criteria, found that ginkgo did not raise the risk for internal bleeding, despite its capacity to increase blood flow. Nevertheless, Roland and Nergard (2012) recommended withdrawing ginkgo two weeks prior to surgery because of its known capacity to increase blood flow.

Kava Kava

Kava kava, or *Piper methysticum*, is a plant known from the islands of the Pacific, where it is used in ritualistic contexts, such as harvesting festivals or group prayers (Chrubasik, 1997). It was also used as a traditional relaxant and has shown at least some potential for anxiolytic and antistress effects (Fugh-Berman & Cott, 1999). In the United States, kava is among the top-selling herbal preparations, with an approximate annual turnover of $8 million (Pittler & Ernst, 2003). In some studies, it has shown anxiolytic and antistress effects (Fugh-Berman & Cott, 1999). Sarris and Kavanagh (2009) reviewed the available evidence and reported that extant

data supported the efficacy of kava only for generalized anxiety. Pittler and Ernst (2003) conducted a systematic review of the efficacy of kava in anxiety disorders and identified twelve trials that met the preset inclusion criteria of being randomized, double-blind, placebo-controlled, and using kava mono-preparations (i.e., preparations that did not contain other ingredients or mixtures of chemical drugs and kava). Seven of these trials reported results suitable for meta-analysis as defined by the authors, that is, a common outcome measure (Hamilton Rating Scale for Anxiety). The weighted mean difference was $d = 3.9$ (CIs 0.1 to 7.7). Nevertheless, the trials were criticized in terms of low sample sizes, failure to describe how randomization was performed, and the method of achieving double-blinding (Pittler & Ernst, 2003). Hence, debate regarding the efficacy of kava for anxiety continues.

A possible mechanism for kava's purported efficacy is not yet known with certitude (Sarris & Kavanagh, 2009). It seems to reduce central noradrenergic activation and modulate GAB-ergic activity, thus influencing arousal and inducing anxiolytic effects. However, kava seems to be another example of a traditional pharmacological substance, the use of which was based only on folklore and unsystematic experience that modern research may ultimately vindicate. Nevertheless, the manufacturing of pharmaceuticals from plant extracts can be a cumbersome process compared with the synthesizing of chemical agents. This may explain why, until recently, the production of phytotherapeutics was the domain of rather small family-owned businesses with little pressure and less funds to carry out research. Only increasing public use has spurred some public funding for research in this area. But before this research could take effect, health authorities issued warnings of suspected liver damage from kava, and so the substance was taken off the market in Canada, the United Kingdom, and the European Union (Escher, Desmeules, Giostra, & Mentha, 2001). Currently, the World Health Organization is reviewing this health warning (Sarris & Kavanagh, 2009), and the FDA still advises about this potential health hazard (Teschke & Schulze, 2010)

Close analysis of the cases shows that the European Union warning was based on seven cases, all of which were dependent on multiple substances such as prescription drugs and alcohol at the same time and had multiple kinds of drug abuse; to attribute liver failure to kava seems to be unsubstantiated and it could be attributed to any of the other drugs in usage and not necessarily reproducible (Schmidt, 2003). The potential problem with kava stems from the fact that it seems to block the activity of the cytochrom P450 enzyme, which is one of the major detoxification enzymes in the human body (Sarris & Kavanagh, 2009). Thus, if kava is taken with other drugs, anxiolytics, or heavy doses of alcohol that rely on CP450 activity, liver toxicity might ensue. Studies of indigenous users show that usage is safe, as long as no other drug interactions have to be considered (Sarris & Kavanagh, 2009).

The remedies mentioned here are those with comparatively good research records. Many other preparations are still used in open, uncontrolled trials. For these preparations, the available evidence is well below what is standard in conventional practice. This fact reflects the situation that until recently research in the entire area of phytotherapy has received no public sponsoring. Because most of the preparations are licensed at least in European countries or in Germany, and many are sold as food or dietary supplements in the United States, there has been no pressure for private sponsors to support research for marketing or licensing reasons. We hope that the public's increasing use of these preparations will engender comparable research efforts that will answer crucial questions regarding their safety and efficacy. At the same time, the public should benefit from awareness of the limits of scientific knowledge regarding herbal remedies, including knowledge of the side effects of these preparations and potentially harmful interactions with conventional prescribed medicines.

PARALLELS BETWEEN ANTIDEPRESSANT MEDICATION AND HERBAL REMEDIES

There is an important parallel in what is known and what remains unknown about conventional antidepressants and complementary and alternative treatments. In both cases, it is difficult to verify specific effects over and above placebo effects in randomized controlled trials. This is not necessarily because of a weak response following these treatments, but rather because of frequent large therapeutic responses to placebo. For example, approximately half of the clinical trials sponsored by the manufacturers of SSRIs have yielded nonsignificant differences between drug and placebo, but have shown substantial clinical improvement in patients no matter whether on placebo or on real drug.

The relatively small differences between active substances and placebos have created the prejudice that complementary and alternative treatments are "nothing but placebos." Yet a similar set of data has left the reputation of conventional antidepressants largely intact. One reason for this divergence may be the difficulty in understanding the possible modes of action and molecular mechanisms of complementary and alternative treatments. Many of the compounds are extremely complex. This makes it difficult to analyze all of their components, let alone to determine which effects they have on which enzymatic or hormone systems. In addition, the compound may have different effects than the individual components or synergistic effects multiplying individual ones. Nevertheless, the mode of action of antidepressants is also not well understood, and it is likely that the divergence in the interpretations of similar data are partly due to ideological reasons as well. Ideological issues aside, scientific research in alternative

medicine is a rather recent enterprise. Researchers have conducted many more trials with chemical substances compared with phytotherapeutics, and the positive data concerning the efficacy and safety of phyotherapeutic substances need to be replicated on a larger scale. Still, there seems to be enough similarity between the two bodies of literature to warrant comparison.

Finally, for both antidepressants and some herbal remedies, the relatively small differences between placebo and active substances do not necessarily mean that the treatments are of little use. If we define usefulness of a treatment only in terms of difference between treatment and placebo in relation to direct and indirect costs of the treatment, versus the costs of the untreated disease, as is implied by the conventional definition of efficacy and utility, then some might conclude that both antidepressants and phytotherapeutic substances are not of much use. However, what matters is not only the relative size of the effect, but also the absolute size compared with baseline, or in other words the magnitude of specific and nonspecific effects combined. This must then be evaluated against the risks and harms associated with these treatments. Although the responses to different treatments for depression are about the same (Khan, Faucett, Lichtenberg, Kirsch, & Brown, 2012), the risks are very different. Antidepressants produce side effects, the most prevalent of which is sexual dysfunction, which is suffered by 70 to 80% of patients given SSRIs. Antidepressants may also increase the risk of suicidal thinking in young people and produce other risks, including death from all causes in the elderly, miscarriages in pregnant women, and autism in the offspring of users (Andrews, Thomson, Amstadter, & Neale, 2012; Domar, Moragianni, Ryley, & Urato, 2013; Rai et al., 2013; Stone et al., 2009). Rather than being a front-line treatment, they should typically be used as a last resort.

Providing a good explanatory myth and preparing a convincing therapeutic ritual are among the common factors of all therapies (Frank, 1987). Hence we can hypothesize that for many people, the potential for nonspecific effects is greater in complementary and alternative treatments. This is particularly true of people who have a worldview compatible with the application of "natural" products and who have a belief system favoring complementary and alternative treatments. For others, who subscribe to a more rational and mechanistic approach to diseases, conventional medical treatments may be more effective, although the risks with which they are associated might preclude their use. For still others, psychotherapy might elicit the greatest expectancy effects, and thereby the greatest therapeutic benefit. Nevertheless, the acid test is whether well-controlled studies bear out this hypothesis. For example, it would be intriguing to determine whether patients requesting a complementary treatment experience stronger positive effects than those who are either opposed or indifferent to such a treatment. Our prediction would be that the difference should be statistically and clinically significant, precisely because the nonspecific

effects can be better harnessed in believers. Indeed, this effect has been demonstrated in a comparison of the use of hypnosis versus nonhypnotic treatment with clients who either did or did not request hypnotic treatment (Lazarus, 1973).

Complementary treatments provide a stimulus that counters either–or thinking in terms of specific versus nonspecific effects. Instead, they invite us to think of specific effects sitting on top of nonspecific effects. In the Middle Ages, modern thinkers were seen as dwarfs sitting on the shoulders of giants and thereby able to see further (Klibansky, 1936). We would like to reinterpret this metaphor: In some cases, specific effects may be dwarfs sitting on the shoulders of nonspecific effects, and this may be the reason that the treatments work so well.

GLOSSARY

Antidepressants: A heterogeneous grouping of different chemical substances, including tricyclics, SSRIs, and MAOIs, that are used for the pharmacological treatment of depression. Antidepressants are considered the first line of intervention in acute and chronic episodes of depression.

Effect size: The primary statistic used in meta-analysis. It is often calculated by dividing the mean difference between a treated group and a control group by the standard deviation of the control group.

Monoamine oxidase inhibitors (MAOIs): Substances that inhibit the enzyme monoamine oxidase (MAO), which degrades serotonin and other monoamines, thereby prolonging the life of these amines and their availability. MAOIs are commonly used as antidepressants, although they are associated with certain severe dietary restrictions.

Phytotherapeutics: Pharmaceutical substances made of plants or extracts from plants. These can be standardized liquid or dried extracts, which guarantee the same amount of active components. In some cases they can be made as decoctions in herbal teas.

Placebo: Of Latin origin, literally, "I will please, I will be a pleasure." In the prepharmaceutical era of the 19th century and the beginning of the 20th century, the term "placebo" was used to denote a treatment with no known efficacy that was given to please a patient. Placebos were later introduced as control substances in trials. In that sense, they are physically (but not necessarily psychologically) inert substances—corn flour, lactose, glycerin, and the like—that are packed, colored, and prepared to resemble the test substance.

Selective serotonin reuptake inhibitors (SSRIs): Since the 1970s, new antidepressants based on the serotonin hypothesis of depression have been developed. They are called selective serotonin reuptake inhibitors (SSRIs) because they inhibit pharmacologically the process by which serotonin released by neurons is taken back into these neurons. The most widely used of these substances is fluoxetine, better known by the brand name Prozac.

Tricyclics: The older antidepressants are known as tricyclics, the most widely used of which are amitryptiline and imipramine, which are standard reference substances in controlled studies of depression.

REFERENCES

Access Economics (2010). *Cost effectiveness of complementary medicine*. Canberra, Australia: National Institute of Complementary Medicine. Retrieved July 30, 2013, from *www.nicm.edu.au/images/stories/research/docs/cost_effectiveness_cm_ae_2010.pdf*.

Ader, R., & Cohen, N. (1991). The influence of conditioning on immune responses. In R. Ader, D. L. Felten, & N. Cohen (Eds.), *Psychoneuroimmunology* (pp. 611–646). San Diego, CA: Academic Press.

Amanzio, M., & Benedetti, F. (1999). Neuropharmacological dissection of placebo analgesia expectation-activated opioid systems versus conditioning-activated specific subsystems. *Journal of Neuroscience*, *19,* 484–494.

Amieva, H., Meillon, C., Helmer, C., Barberger-Gateau, P., & Dartigues, J. F. (2013). *Ginkgo biloba* extract and long-term cognitive decline: A 20-year follow-up population based study. *PLoS ONE, 8*(1), e52755.

Andrews, P. W., Thomson, J. A., Amstadter, A., & Neale, M. C. (2012). *Primum non nocere*: An evolutionary analysis of whether antidepressants do more harm than good [Review]. *Frontiers in Psychology, 3*.

Astin, J. A. (1998). Why patients use alternative medicine: Results of a national study. *Journal of the American Medical Association*, *279,* 1548–1553.

Baethge, C., Assall, O. P., & Baldessarini, R. J. (2013). Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000–2010. *Psychotherapy and Psychosomatics, 82*(3), 152–160.

Barnes, P. M., Bloom, B., & Nahin, R. L. (2008). *Complementary and alternative medicine use among adults and children: United States, 2007.* Hyattsville, MD: National Center for Health and Statistics.

Barrett, B., Vohmann, M., & Calabrese, C. (1999). *Echinacea* for upper respiratory infection. *Journal of Family Practice*, *48,* 628–635.

Benedetti, F., Amanzio, M., Baldi, S., Casadio, C., & Maggi, G. (1999). Inducing placebo respiratory depressant responses in humans via opioid receptors. *European Journal of Neuroscience*, *11,* 625–631.

Berner, M. (2001). Kommentar zu Shelton, et al. (2001): Hypericum bei Depression [Commentary on Shelton et al. (2001): Hypericum in depression]. *Forschende Komplementärmedizin/Research in Complementary and Classical Natural Medicine, 8*.

Birks, J., & Grimley Evans, J. (2009). *Ginkgo biloba* for cognitive impairment and dementia [Review]. *Cochrane Database of Systematic Reviews*, *1*, CD003120.

Blashki, T. G., Mowbray, R., & Davies, B. (1971). Controlled trial of amitryptiline in general practice. *British Medical Journal, 1*(741), 133–138.

Bundesinstitut für Arzneimittel und Medizinprodukte. (2000). Bekanntmachung über die Registrierung, Zulassung und Nachzulassung von Arzneimitteln: Abwehr von Arzneimittelrisiken, Anhörung, Stufe II: Johanniskrauthaltige (Hypericum) Humanarzneimittel zur innerlichen Anwendung vom 24.März 2000 [Notice on registration, and admission of medicines: Defense of risks from medicines, hearing #2: Hypericum containing drugs for humans as inward applications, March 24, 2000]. *Bundesanzeiger*, *52,* 6009–6010.

Centers for Disease Control and Prevention. (2011, March 31). An estimated 1 in 10 U.S. adults report depression. Retrieved August 1, 2013, from *www.cdc.gov/features/dsdepression*.

Chan, P.-C., Xia, Q., & Fu, P. P. (2007). *Ginkgo biloba* leave extract: Biological, medicinal, and toxicological effects. *Journal of Environmental Science and Health Part C, 25*, 211–244.

Chrubasik, S. (1997). Klinisch geprüfte Wirksamkeit bei nervösen Angst-, Spannungs- und Unruhezuständen [Clinically proven efficacy in nervous anxiety, tension and states of unrest]. *Der Allgemeinarzt*, *18,* 1683–1687.

Chrubasik, S., Junck, H., Zappe, H., & Stutzke, O. (1998). A survey on pain complaints and health care utilization in a German population sample. *European Journal of Anaesthesiology*, *15,* 397–408.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cordell, G. A. (2002). Natural products in drug discovery—creating a new vision. *Phytochemistry Reviews, 1*, 261–273.

Dawes, R. M. (1998). Commentary on Kirsch and Saperstein. *Prevention and Treatment, 1,* Article 0005c. Available at *http://journals.apa.org/prevention/volume1/pre0010005c.html*.

DeKosky, S. T., Williamson, J. D., Fitzpatrick, A. L., Kronmal, R. A., Ives, D. G., Saxton, J. A., et al. (2008). *Ginkgo biloba* for prevention of dementia: A randomized controlled trial. *Journal of the American Medical Association, 300*, 2253–2262.

Domar, A. D., Moragianni, V. A., Ryley, D. A., & Urato, A. C. (2013). The risks of selective serotonin reuptake inhibitor use in infertile women: A review of the impact on fertility, pregnancy, neonatal health and beyond. *Human Reproduction, 28*(1), 160–171.

Eardly, S., Bishop, F. L., Prescott, P., Cardini, F., Brinkhaus, B., Santos-Rey, K., et al. (2012). A systematic literature review of complementary and alternative medicine prevalence in EU. *Forschende Komplementärmedizin, 19*(Suppl. 2), 18–28.

Easthope, G., Tranter, B., & Gill, G. (2001). General practitioners' attitudes toward complementary therapies. *Social Science and Medicine, 51*, 1555–1561.

Escher, M., Desmeules, J., Giostra, E., & Mentha, G. (2001). Hepatitis associated with kava, a herbal remedy for anxiety. *British Medical Journal, 322*, 139.

Field, H. L., Monti, D. A., Greeson, J. M., & Kunkel, E. J. S. (2000). St. John's wort. *International Journal of Psychiatry in Medicine*, *30,* 203–219.

Fisher, S., & Greenberg, R. P. (Eds.). (1989). *The limits of biological treatments for psychological distress*. Hillsdale, NJ: Erlbaum.

Fisher, S., & Greenberg, R. P. (1993). How sound is the double-blind design for evaluating psychotropic drugs? *Journal of Nervous and Mental Diseases, 181*, 345–350.

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., et al. (2010). Antidepressant drug effect and depression severity. *Journal of the American Medical Association, 303*, 47–53.

Frank, J. D. (1987). Therapeutic components shared by all psychotherapies. In J. H. Harvey & M. M. Parks (Eds.), *Psychotherapy research and behavior change* (pp. 73–122). Washington, DC: American Psychological Association.

Fugh-Berman, A., & Cott, J. (1999). Dietary supplements and natural products as psychotherapeutic agents. *Psychosomatic Medicine*, *61,* 712–728.

Furnham, A., & Kirkcaldy, B. (1996). The health beliefs and behaviours of orthodox and complementary medicine clients. *British Journal of Clinical Psychology*, *35,* 49–61.

Gerson, S., Belin, T. R., Kaufman, A., Mintz, J., & Jarvik, L. (1999). Pharmacological and psychological treatments for depressed older patients: A meta-analysis and overview of recent findings. *Harvard Review of Psychiatry, 7*, 1–28.

Greenberg, R. P., Bornstein, R. F., Zborowske, M. J., Fisher, S., & Greenberg, M. D. (1994). A meta-analysis of fluoxetine outcome in treatment of depression. *Journal of Nervous and Mental Diseases, 182*, 547–551.

Greeson, J. M., Sanford, B., & Monti, D. A. (2001). St. John's wort (*Hypericum perforatum*): A review of the current pharmacological, toxicological, and clinical literature. *Psychopharmacology*, *153*, 402–414.

Gundling, K., & Ernst, E. (1999). Complementary and alternative medicine in cardiovascular disease: What is the evidence it works? *Western Journal of Medicine*, *171*, 191–194.

Ihl, R. (2012). *Ginkgo biloba* extract EGb 761: Clinical data in dementia. *International Psychogeriatrics, 24*(Suppl. 1), S35–S40.

Joffe, R., Sokolov, S., & Streiner, D. (1996). Antidepressant treatment of depression: A meta-analysis. *Canadian Journal of Psychiatry, 41*, 613–616.

Kaschel, R. (2009). *Ginkgo biloba*: Specificity of neuropsychological improvement—a selective review in search of differential effects. *Human Psychopharmacology: Clinical and Experimental, 24*, 345–370.

Kaschel, R. (2011). Specific memory effects of *Ginkgo biloba* extract EGb 761 in middle-aged healthy volunteers. *Phytomedicine, 18*, 1202–1207

Kasper, S., Gastpar, M., Möller, H.-J., Müller, W. E., Volz, H.-P., Dienel, A., et al. (2010). Better tolerability of St. John's wort extract WS 5570 compared to treatment with SSRIs: A reanalysis of data from controlled clinical trials in acute major depression. *International Clinical Psychopharmacology, 25*, 204–213.

Kellermann, A. J., & Kloft, C. (2011). Is there a risk of bleeding associated with standardized *Ginkgo biloba* extract therapy?: A systematic review and meta-analysis. *Pharmacotherapy, 31*, 490–502.

Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I., & Brown, W. A. (2012). A systematic review of comparative efficacy of treatments and controls for depression. *PLoS ONE, 7*(7), e41778.

Khan, A., Warner, H. A., & Brown, W. A. (2000). Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: An analysis of the Food and Drug Administration database. *Archives of General Psychiatry 57,* 311–317.

Kirsch, I. (1998). Reducing noise and hearing placebo more clearly. *Prevention and Treatment, Article 0007r.* Available at *http://journals.apa.org/prevention/volume1/pre0010007a.html.*

Kirsch, I. (2000). Are drug and placebo effects in depression additive? *Biological Psychiatry, 47,* 733–735.

Kirsch, I. (2010). *The emperor's new drugs: Exploding the antidepressant myth.* New York: Basic Books.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine, 5*(2). Available at *www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0050045.*

Kirsch, I., Moore, T. J., Scoboria, A., & Nicholls, S. S. (2002). The emperor's new

drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention and Treatment, 5*(1), 23.

Kirsch, I., & Sapirstein, G. (1998). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention and Treatment, 1, Article 0002a*. Available at *www.journals.apa.org/prevention/volume1/pre0010002a.html*.

Kirsch, I., & Sapirstein, G. (1999). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medications. In I. Kirsch (Ed.), *Expectancy, experience, and behavior* (pp. 303–320). Washington, DC: American Psychological Association.

Kleijnen, J., & Knipschild, P. (1992a). *Ginkgo biloba* for cerebral insufficiency. *British Journal of Clinical Pharmacology, 34*, 352–358.

Kleijnen, J., & Knipschild, P. (1992b). *Ginkgo biloba*. *Lancet, 340*, 1136–1139.

Klein, D. F. (1998). Listening to meta-analysis but hearing bias. *Prevention and Treatment, 1,* Article 0006c. Available at *http://journals.apa.org/prevention/volume1/pre0010006c.html*.

Klibansky, R. (1936). Standing on the shoulders of the giants. *Isis*, *26,* 147–149.

Kuller, L. H., Ives, D. G., Fitzpatrick, A. L., Carlson, M. C., Mercado, C., Lopez, O. L., et al. (2010). Does *Ginkgo biloba* reduce the risk of cardiovascular events? *Circulation: Cardiovascular Quality and Outcomes, 3*, 41–47.

Laakmann, G., Schüle, C., Baghai, T., & Kieser, M. (1998). St. John's wort in mild to moderate depression: The relevance of hyperforin for the clinical efficacy. *Pharmacopsychiatry*, *31,* 54–59.

LaFrance, W. C., Lauterbach, E. C., Coffey, C. E., Salloway, S. P., Kaufer, D. I., Reeve, A., et al. (2000). The use of herbal alternative medicines in neuropsychiatry: A report of the ANPA Committee on Research. *Journal of Neuropsychiatry and Clinical Neurosciences*, *12,* 177–192.

Laws, K. R., Sweetnam, H., & Kondel, T. K. (2012). Is *Ginkgo biloba* a cognitive enhancer in healthy individuals?: A meta-analysis. *Human Psychopharmacology: Clinical and Experimental, 27*, 527–533.

Lazarus, A. A. (1973). "Hypnosis" as a facilitator in behavior therapy. *International Journal of Clinical and Experimental Hypnosis, 21,* 25–31.

Linde, K., Berner, M., Egger, M., & Mulrow, C. D. (2005). St. John's wort for depression: Meta-analysis of randomised controlled trials. *British Journal of Psychiatry, 186*, 99–107.

Linde, K., Berner, M. M., & Kriston, L. (2008). St. John's wort for major depression. *Cochrane Database of Systematic Reviews*, *4*, Art. No. CD000448.

Linde, K., & Mulrow, C. D. (2001). St. John's wort for depression (Cochrane Review). *Cochrane Library*, *3*. Oxford, UK: Update Software.

Linde, K., Ramirez, G., Mulrow, C. D., Pauls, A., Weidenhammer, W., & Melchart, D. (1996). St John's wort for depression—an overview and meta-analysis of randomised clinical trials. *British Medical Journal*, *313,* 253–258.

Loef, M., & Walach, H. (2012). The combined effects of healthy lifestyle behaviors on all cause mortality: A systematic review and meta-analysis. *Preventive Medicine, 55*, 163–170.

Madden, K. S., Boehm, G. W., Lee, S. C., Grota, L. J., Cohen, N., & Ader, R. (2001). One-trial conditioning of the antibody response to hen egg lysozyme in rats. *Journal of Neuroimmunology*, *113,* 236–239.

Marstedt, G., & Moebus, S. (2002). Inanspruchnahme alternativer Methoden in

der Medizin [Usage of alternative methods in medicine]. *Gesundheitsberichterstattung des Bundes* (Heft 9). Retrieved July 30, 2013, from *www.gbe-bund.de/gbe10/ergebnisse.prc_tab?fid=7861&suchstring=naturheilverfahren&query_id=&sprache=D&fund_typ=TXT&methode=2&vt=1&verwandte=1&page_ret=0&seite=&p_lfd_nr=7&p_news=&p_sprachkz=D&p_uid=gast&p_aid=3589587&hlp_nr=3&p_janein=J#SEARCH=%2522naturheilverfahren%2522*.

Melchart, D., Linde, K., Worku, F., Bauer, R., & Wagner, H. (1994). Immunmodulation mit *Echinacea*-haltigen Arzneimitteln [Immuno-modulation with drugs containing *Echinacea*]. *Forschende Komplementärmedizin*, *1,* 27–36.

Melchart, D., Linde, K., Worku, F., Sarkady, L., Holzmann, M., Jurcic, K., et al. (1995). Results of five randomized studies on the immunmodulatory activity of preparations of *Echinacea*. *Journal of Alternative and Complementary Medicine*, *1,* 145–160.

Melchart, D., Walther, E., Linde, K., Brandmaier, R., & Lersch, C. (1998). *Echinacea* root extracts for the prevention of upper respiratory tract infections: A double-blind, placebo-controlled randomized trial. *Archives of Family Medicine*, *7,* 541–545.

Mills, S. (2001) Herbal medicine. In G. Lewith, W. B. Jonas, & H. Walach (Eds.), *Clinical research for complementary medicine*: *Principles, problems, solutions*. London: Churchill Livingstone.

Moncrieff, J., & Kirsch, I. (2005). Efficacy of antidepressants in adults. *British Medical Journal, 331*, 155–157.

Mössner, R., & Lesch, K.-P. (1998). Role of serotonin in the immune system and in neuroimmune interactions. *Brain, Behavior, and Immunity*, *12,* 249–271.

NICE (National Institute for Clinical Excellence). (2004). Depression: Management of depression in primary and secondary care. *Clinical Practice Guideline* No 23. Retrieved May 24, 2005, from *www.nice.org.uk/page.aspx?o=235213*.

Nicolai, S. P. A., Kruidenier, L. M., Bendermacher, B. L. W., Prins, M. H., Stokmans, R. A., Broos, P. P. H. L., et al. (2013). *Ginkgo biloba* for intermittent claudication. *Cochrane Database of Systematic Reviews,* published online.

Nissen, N., Schunder-Tatzber, S., Weidenhammer, W., & Johannessen, H. (2012). What attitudes and needs do citizens in Europe have in relation to CAM? *Forschende Komplementärmedizin, 19*(Suppl. 2), 9–17.

Oken, B. S, Storzbach, D. M. & Kaye, J. A. (1998). The efficacy of *Ginkgo biloba* on cognitive function in Alzheimer disease. *Archives of Neurology, 55*, 1409–1415.

Ott, B. R., & Owens, N. J. (1998). Complementary and alternative medicines for Alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology*, *11*, 163–173.

Philipp, M., Kohnen, R., & Hiller, K. O. (1999). Hypericum extract versus imipramine or placebo in patients with moderate depression: Randomised multicentre study of treatment for eight weeks. *British Medical Journal, 319,* 1534–1539.

Pigott, H. E., Leventhal, A. M., Alter, G. S., & Boren, J. J. (2010). Efficacy and effectiveness of antidepressants: Current status of research. *Psychotherapy and Psychosomatics, 79*, 267–279.

Pincus, D., & Walach, H. (Eds.). (2012). *Complexity, whole systems research, and complementary therapies: New concepts and methods*. Basel, Switzerland: Karger.

Pirotta, M. V., Cohen, M. M., Kotsirilos, V., & Farish, S. J. (2000). Complementary therapies: Have they become accepted in general practice? *Medical Journal of Australia*, *172,* 105–109.

Pittler, M. H., & Ernst, E. (2003). Kava extract versus placebo for treating anxiety. *Cochrane Database of Systematic Reviews*, *1*, Art. No. CD003383.

Pratkanis, A. R. (1995, July/August). How to sell a pseudoscience. *Skeptical Inquirer, 19*, 19–25.

Pratt, L. A., Brody, D. J., & Gu, Q. (2011). *Antidepressant use in persons aged 12 and over: United States, 2005–2008*. Washington, DC: National Center for Health Statistics. Retrieved from *www.cdc.gov/nchs/data/databriefs/db76.pdf*.

Quitkin, F. M., Rabkin, J. G., Gerald, J., Davis, J. M., & Klein, D. F. (2000). Validity of clinical trials of antidepressants. *American Journal of Psychiatry, 157,* 327–337.

Rabkin, J. G., Markowitz, J. S., Stewart, J. W., McGrath, P. J., Harrison, W., Quitkin, F. M., et al. (1986). How blind is blind?: Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. *Psychiatry Research, 19*, 75–86.

Rahimi, R., Nikfar, S., & Abdollahi, M. (2009). Efficacy and tolerability of *Hypericum perforatum* in major depressive disorder in comparison with selective serotonin reuptake inhibitors: A meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 33*, 118–127.

Rai, D., Lee, B. K., Dalman, C., Golding, J., Lewis, G., & Magnusson, C. (2013). Parental depression, maternal antidepressant use during pregnancy, and risk of autism spectrum disorders: Population based case-control study. *British Medical Journal, 346*.

Rapaport, M. H., Nierenberg, A. A., Howland, R., Dording, C., Schettler, P. J., & Mischoulon, D. (2011). The treatment of minor depression with St. John's wort or citalopram: Failure to show benefit over placebo. *Journal of Psychiatric Research, 45*, 931–941.

Roland, P. D., & Nergard, C. S. (2012). *Ginkgo biloba*: Effect, adverse events and drug interactions. *Tidsskrift for Den Norske Laegeforening, 132*, 956–959.

Rosslenbroich, B., & Saller, R. (1992) Phytotherapie im Überblick [A review on phytotherapeutics]. In M. Bühring & F. H. Kemper (Eds.), *Naturheilverfahren und unkonventionelle medizinische Richtungen [Naturopathic therapies and unconventional medical practices*]. Berlin: SpringerLoseblattSysteme, Section 8.01.

Rush, J. A., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., et al. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *American Journal of Psychiatry, 163*, 1905–1917.

Sarris, J., & Kavanagh, D. J. (2009). Kava and St. John's wort: Current evidence for use in mood and anxiety disorders. *Journal of Alternative and Complementary Medicine, 15*, 827–836.

Schmidt, M. (2003). Are kavalactones the hepatotoxic principle of kava extracts?: The pitfalls of the gluthathione theory. *Journal of Alternative and Complementary Medicine, 9*, 183–188.

Schonauer, K. (1994). *Semiotic foundation of drug therapy: The placebo problem in a new perspective*. Berlin: Mouton de Gruyter.

Schulz, V. (2000). The psychodynamic and pharmacodynamic effects of drugs: A differentiated evaluation of the efficacy of phytotherapy. *Phytomedicine*, *7,* 73–81.

Shelton, R. C., Keller, M. B., Gelenberg, A., Dunner, D. L., Hirschfeld, R., Thase, M. E., et al. (2001). Effectiveness of St. John's wort in major depression: A randomized controlled trial. *Journal of the American Medical Association, 285*, 1978–1986.

Singer, A., Wonnemann, M., & Müller, W. E. (1999). Hyperforin, a major antidepressant constituent of St. John's wort, inhibits serotonin uptake by elevating free intracelluclar Na+. *Journal of Pharmacology and Experimental Therapeutics, 290*, 1363–1368.

Snitz, B. E., O'Meara, E. S., Carlson, M. C., Arnold, A. M., Ives, D. G., Rapp, S. R., et al. (2009). *Ginkgo biloba* for preventing cognitive decline in older adults. *Journal of the American Medical Association, 302*, 2663–2670.

Solomon, D., Adams, J., & Graves, N. (2013). Economic evaluation of St. John's wort (*Hypericum perforatum*) for the treatment of mild to moderate depression. *Journal of Affective Disorders, 148*, 228–234.

Sparber, A., Wootton, J. C., Bauer, L., Curt, G., Eisenberg, D., Levin, T., et al. (2000). Use of complementary medicine by adult patients participating in HIV/AIDS clinical trials. *Journal of Alternative and Complementary Medicine*, *5,* 415–422.

Stone, M., Laughren, T., Jones, M. L., Levenson, M., Holland, P. C., Hughes, A., et al. (2009). Risk of suicidality in clinical trials of antidepressants in adults: Analysis of proprietary data submitted to U.S. Food and Drug Administration. *British Medical Journal, 339.*

Teschke, R., & Schulze, R. (2010). Risk of kava hepatotoxicity and the FDA consumer advisory. *Journal of the American Medical Association, 304*(19), 2174–2175.

Wagner, H., & Jurcic, K. (1991). Immunologische Untersuchungen von pflanzlichen Kombinationspräparaten [Immunological investigations of plant combination drugs]. *Arzneimittelforschung/Drug Research*, *41*(2), 1072–1076.

Walach, H., & Güthlin, C. (2000). Effects of acupuncture and homeopathy: A prospective documentation. Intermediate results. *British Homeopathic Journal*, *89,* 31–34.

Walach, H., & Maidhof, C. (1999). Is the placebo effect dependent on time? In I. Kirsch (Ed.), *Expectancy, experience, and behavior* (pp. 321–332). Washington, DC: American Psychological Association.

Walach, H., Sadaghiani, C., Dehm, C., & Bierman, D. J. (2005). The therapeutic effect of clinical trials: Understanding placebo response rates in clinical trials—A secondary analysis. *Medical Research Methodology, 5*, 26.

Wisniewski, S., Rush, A., Nierenberg, A., Gaynes, B., Warden, D., Luther, J., et al. (2009). Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR* D report. *American Journal of Psychiatry, 166*(5), 599–607.

Woelk, H. for the Remotiv/Imipramine Study Group. (2000). Comparison of St. John's wort and imipramine for treating depression: Randomised controlled trial. *British Medical Journal, 321,* 536–539.

PART IV

CONTROVERSIES IN THE TREATMENT OF CHILD AND ADOLESCENT DISORDERS

CHAPTER THIRTEEN

Empirically Supported, Promising, and Unsupported Treatments for Attention-Deficit/ Hyperactivity Disorder

Daniel A. Waschbusch
and James G. Waxmonsky

As operationalized in DSM-5, attention-deficit/hyperactivity disorder (ADHD) is a chronic disorder characterized by abnormally high levels of inattention, impulsivity, and/or overactivity (American Psychiatric Association, 2013). Although ADHD is found across the lifespan (Sibley et al., 2012a, 2012b), it is most often diagnosed in elementary school-age children. ADHD is estimated to affect 5% of elementary school-age children worldwide (Polanczyk & Rohde, 2007), although estimates in the United States have ranged as high as 11% (Schwarz & Cohen, 2013; Visser, Bitsko, Danielson, & Perou, 2010; Wolraich et al., 2012). ADHD is a common reason for referral to mental health clinics (Aupont et al., 2012; Thompson & Ni Bhrolchain, 2011) and is associated with serious impairment in many domains of daily functioning, including peer rejection, negative interactions with parents, classroom behavior problems, and academic difficulties (Barkley, 2005). ADHD in childhood is associated with serious problems in adolescence and adulthood, including school and vocational failure (Kent et al., 2011; Kuriyan et al., 2012), disruptions in interpersonal relationships (Bagwell, Molina, Pelham, & Hoza, 2001; Wymbs et al., 2011), criminal behavior (Sibley et al., 2011), mental health problems (Hechtman, 2006), and alcohol or other substance abuse (Molina, Pelham, Gnagy, Thompson,

& Marshal, 2007). The most severe outcomes are often associated with the co-occurrence of ADHD and conduct problems (Waschbusch, 2002). The costs of ADHD are substantial for children, families, schools, and communities (Pelham, Foster, & Robb, 2007; Robb et al., 2011). Thus, identification of effective treatments for ADHD has major public health implications.

A wide variety of treatments have been proposed for ADHD, perhaps more so than for any other form of child psychopathology. Purported treatments range from medication to restrictive or supplemental diets, allergy treatments, biofeedback, homeopathic remedies, and many others. The wide variety and prevalence of both well-supported treatments and pseudotreatments raises questions as to which treatments for ADHD work and which do not, and how to tell the difference between them. In this chapter, we will address these questions by selectively reviewing the research evidence. We first review treatments that have a solid evidence base and are considered empirically supported. Second, we review promising treatments for ADHD. These are treatments that have been investigated using methodologically sound research and shown to produce positive treatment response in at least some studies, although the body of evidence is not sufficient to consider them empirically supported. This tentative verdict may be due to a limited number of studies or to inconsistent findings across studies. Third, we review unsupported treatments for ADHD. These are treatments that have been researched and shown to be contraindicated for ADHD. We believe that grouping treatments into empirically supported, promising, and unsupported categories is a useful heuristic for organizing treatments, but we hasten to add that placement of treatments into these categories should be considered dynamic rather than static. New research may cement a treatment into its current category or shift it to another. We encourage readers to consume new research on these treatments and incorporate it into their own evaluations.

EMPIRICALLY SUPPORTED TREATMENTS

Over the past two decades a concerted effort has been made to identify empirically supported treatments (ESTs) for use with children and adolescents. In the 1990s, the Clinical Child Psychology section of the Clinical Psychology division of the American Psychological Association formed a task force to identify ESTs for children and adolescents (Lonigan, Elbert, & Johnson, 1998). The task force developed criteria for evaluating the evidence base, and these were published in 1998 in a special issue of the *Journal of Clinical Child Psychology*, which has since been renamed the *Journal of Clinical Child and Adolescent Psychology*. Treatments were considered to be empirically supported if research showed they were superior to placebo treatments in either of two ways: (1) using at least two group-design studies

conducted by different sets of investigators, or (2) using at least 10 single-case studies. In addition, the treatments must have been manualized, and the samples used to investigate the treatment must have been clearly specified. These criteria were then applied by reviewing the evidence base for several mental health disorders experienced by children, including ADHD (Pelham, Wheeler, & Chronis, 1998). This original review was updated a decade later to incorporate new research (Pelham & Fabiano, 2008).

Both the original and more recent review concluded that currently three treatments for ADHD are empirically supported: (1) behavior therapy; (2) central nervous system stimulant medications; and (3) the combination of these two. Arguments for and against each of these treatments were well articulated in the reviews already cited, so we summarize this information only briefly here and focus on presenting new information that has emerged since the publication of these reviews. Specifically, we discuss advances in medication treatments for ADHD, including the introduction of extended-release forms of stimulant medication and the introduction of nonstimulant medications. These interventions have emerged in recent years and now meet the criteria as empirically supported treatments for ADHD. We end this section with a brief review of the Multimodal Treatment of ADHD (MTA) study, as the results are directly relevant to the topic of empirically supported treatment of ADHD and the need for additional treatment approaches.

Stimulant Medications

Central nervous system (CNS) stimulants are well established treatments for pediatric ADHD, with thousands of studies dating back more than 50 years that demonstrate their therapeutic effects (Greenhill et al., 2002). Stimulant medications typically produce significant improvements in the core symptoms of ADHD (see Swanson, McBurnett, Christian, & Wigal, 1995, for a review). There is also evidence that stimulants produce short-term improvements in other associated features of ADHD, such as social behaviors (Pelham, Waschbusch, Hoza, Pillow, & Gnagy, 2001), academic productivity (Pelham et al., 2002), and classroom behavior (Pelham et al., 1993). These changes are seen across multiple measures, including parent and teacher ratings, observational measures, and performance measures (e.g., attention and inhibitory control tasks). Furthermore, these improvements are often immediate and meaningful, and frequently accompanied by few side effects.

Given these findings, it is perhaps not surprising that stimulant medication appears to be the most common form of treatment for ADHD (Bird et al., 2008). The number of prescriptions for ADHD medications has increased over the past decade, with nearly 5% of the general population of children and adolescents having used them at some time in their youth (Visser et al., 2010). ADHD medications have now surpassed antibiotics

as the most commonly prescribed medication for adolescents (Chai et al., 2012) and are the second most commonly prescribed medication class in children.

Two classes of stimulants are used to treat ADHD: methylphenidate (MPH) derivatives and amphetamine (AMPH) salt derivatives. Prior to the 1990s, stimulant medications for ADHD were exclusively short-acting/immediate-release preparations. This state of affairs changed with the introduction of extended-release formulations of both amphetamine salts (such as MAS XR, trade name: Adderall XR) and MPH (such as OROS-MPH, trade name: Concerta). The emergence of these medications alleviated the need for in-school dosing by reliably extending medication effects across the school day and into the afternoon. The trend of developing new formulations of stimulant medication has continued in recent years, and several new products have come to market that are now widely used.

One such product is lisdexamfetamine (trade name: Vyvanse). Vyvanse is an extended-release dextroamphetamine product—that is, it has the same active chemical moiety as dexedrine. Vyvanse is a prodrug, meaning that the active ingredient (dextroamphetamine) becomes bioavailable only after oral ingestion. Essentially, no active medication is released unless the pill is ingested, making it harder to misuse. The prodrug mechanism appears to reduce the risk for abuse, with studies showing that individuals find it less desirable for misuse than older pill forms of stimulants (Jasinski & Krishnan, 2009). However, it is still classified as a controlled substance, so a potential for misuse, while reduced, may exist. Vyvanse has been approved for the treatment of ADHD in children ages 6 and up (as well as adults), with approved doses that range from 20 mg to 70 mg, and with 50 mg being the most common endpoint dose. It has now been extensively researched in several controlled trials that collectively enrolled over 1,000 children.

In addition to a reduced abuse liability, the prodrug nature of Vyvanse appears to produce more predictable blood levels across the day than prior extended-release versions of AMPH (Ermer et al., 2010). Controlled studies of Vyvanse with children and adults show that therapeutic effects are evident 1.5 hours after dosing and continue 13 to 14 hours (S. B. Wigal, Kollins, Childress, Squires, & 311 Study Group, 2009; T. Wigal et al., 2010). Hence, a dose given at 8 a.m. appears to produce effects into the evening, providing therapeutic benefit to older children and adolescents who often do homework or other attention-demanding tasks (such as driving) late into the evening. Like other stimulants, the most commonly reported side effects associated with Vyvanse are appetite suppression, weight loss, headache, irritability, and insomnia. However, because of its extended therapeutic duration, Vyvanse may produce greater appetite suppression and a higher likelihood of insomnia compared with other shorter-acting stimulant preparations (Childress & Sallee, 2012; Faraone, Spencer, Kollins, & Glatt, 2010). If medication is to be considered, the American Academy of

Pediatrics has recently recommended that methylphenidate products be considered the initial medication choice for ADHD in children under the age of 6 (American Academy of Pediatrics, 2011).

The issues of drug "rebound," presenting as worse ADHD symptoms than before medication was used, has been hypothesized to occur with stimulant medications, especially those with a short therapeutic duration. There has been little formal investigation of this topic, but one recent study found little evidence of rebound effects in response to Vyvanse (Lopez et al., 2013). In a post hoc analysis, parents reported that worsening mood at the end of day—when medication would typically wear off and rebound would be expected to occur—was no more likely with active medication than with placebo (Lopez et al., 2013). Only 3% of children prescribed Vyvanse met criteria for rebound, with baseline emotional lability predicting its occurrence. These results suggest that emotional rebound is uncommon and more likely due to return of baseline mood lability as the medication fades rather than a drug-induced adverse event.

A second extended-release stimulant formulation that has come to market in the last decade is dexmethylphenidate (DMPH) tablets and extended-release capsules (trade name: Focalin and Focalin XR). Most medications consist of enantiomers, or left and right halves that are mirror images (like a pair of gloves) and that fit together to form the active agent. In many cases, the pharmacological activity lies primarily in one of the enantiomers. In the case of MPH, it is the *d*-enantiomer that produces most to all of the CNS effects. It has been theorized that a single enantiomer version containing only the *d* half (DMPH) may produce better results or tolerability. While numerous controlled trials have found that both the tablet and extended-release capsule versions of DMPH reduce ADHD symptoms relative to placebo, distinct advantages over traditional MPH products have not been clearly demonstrated (Moen & Keam, 2009). DMPH products are still classified as controlled substance medications and appear to have relatively comparable side effect profiles to MPH. Hence, Focalin and Focalin XR are viable additions to the ADHD pharmacotherapy regiment, but there is no clear evidence that they should be preferentially prescribed over other MPH products.

A third extended-release stimulant treatment option for ADHD that has come to market over the past decade is transdermal methylphenidate (trade name: Daytrana). As the primary market for ADHD medications remains school-aged children, an appreciable number of patients have difficulty swallowing capsulated versions of these medications. Pill-swallowing exercises have proven helpful for only some patients (Beck, Cataldo, Slifer, Pulbrook, & Guhman, 2005). Using a skin patch to deliver stimulant medication is a viable solution for many of these children. An additional potential advantage of transdermal application is the ability to flex the therapeutic duration of the medication by removing the patch at different times of the day. For example, if a child has more homework, the patch can be left

on longer that day. This flexibility also creates the potential for the child to remove the medication earlier than the parent desires, which could be problematic in youth with a high level of oppositional behaviors. Double-blind placebo controlled studies show that the patch has comparable therapeutic and tolerability profiles as oral versions of MPH (Findling et al., 2008; Patrick, Straughn, Perkins, & González, 2009), with the exception of the need to check for skin irritation and sensitization. The maximum recommended wear time for the methylphenidate patch is 9 hours, with therapeutic effects lasting 2 to 4 hours after removal (Wilens et al., 2008). Because only one-third of the medication in the patch is depleted after nine hours of wear time (Patrick et al., 2009), failure to remove the patch could lead to increased side effects later into the evening.

Having greater choice in types, doses, formulations, and delivery systems is clearly beneficial to individuals with ADHD; however, there is little evidence of reliable predictors of treatment effects that can identify the best medication for an individual patient. As a consequence, the process of finding the best stimulant medication for an individual patient can still entail a fair amount of trial and error. Moreover, despite the large base of empirical support for stimulant medications in the treatment of ADHD, there are limits to this treatment. First, treatment gains are maintained only as long as the child is actively taking the medication. Second, not all children with ADHD respond favorably to stimulant medication. Some estimates suggest that about 70–85% of children with ADHD show a positive response, but the remaining 15–30% show either an adverse response or no response. Arguably the most important limitation is that the short-term improvements associated with stimulant medications do not appear to translate into better long-term outcomes (e.g., Molina et al., 2009). These critiques are not unique to stimulant medication; most if not all are shared by other treatments for ADHD.

Nonstimulant Medications

In 2002, atomoxetine (trade name: Strattera) became the first nonstimulant medication approved by the U.S. Food and Drug Administration (FDA) for the treatment of ADHD in children. Sufficient evidence has emerged over the ensuing decade that it can now be considered an EST for ADHD. Strattera is a selective noradrenergic reuptake inhibitor that leads to dopamine elevation in the prefrontal cortex but not elsewhere. The full therapeutic effects of a dose are typically not apparent until up to 4 weeks after initiation (Newcorn et al., 2008). Strattera should be taken every day for optimal efficacy. Once a day dosing may lead to full day effects, making morning or nighttime dosing possible (Waxmonsky, Waschbusch, Akinnusi, & Pelham, 2011; Wehmeier, Dittmann, Schacht, Helsberg, & Lehmkuhl, 2009).

Over 100 controlled trials have documented the efficacy of Strattera for improving ADHD symptoms for individuals as young as 6 years

old. Follow-up studies have documented that the vast majority of patients maintain full response for a year or more (Buitelaar et al., 2007; Wilens et al., 2006). Strattera appears to be equally effective for inattention versus hyperactivity/impulsivity and may be beneficial to several comorbid conditions, including oppositional behaviors (Biederman et al., 2007), anxiety (Geller et al., 2007), and enuresis (Sumner, Schuh, Sutton, Lipetz, & Kelsey, 2006). However, Strattera appears to be somewhat less efficacious than stimulants for ADHD symptoms. Specifically, on measures of symptom reduction, Strattera typically produces effect sizes (Cohen's *d*'s) of about 0.60, whereas stimulants typically produce effect sizes of 0.80 or higher on the same measures (Garnock-Jones & Keating, 2009). Further, about 40% of individuals are nonresponsive to Strattera, a rate substantially higher than that seen with stimulants (Newcorn, Sutton, Weiss, & Sumner, 2009). Given these findings, it is perhaps better to evaluate Strattera as an alternative to, rather than a competitor with, stimulants. Indeed, over 40% of individuals nonresponsive to stimulants may respond to Strattera, and individuals who are nonresponsive to Strattera may respond to stimulant medication (Newcorn et al., 2008).

Side effects typically associated with Strattera include headache, sedation, nausea, and irritability, the last-named occurring more commonly in younger children (Garnock-Jones & Keating, 2009). Strattera is not a controlled substance medication, so there is little risk for abuse (Heil et al., 2002), and it is not associated with suppression of weight, height, or sleep (Spencer et al., 2005). Rare cases of liver toxicity have been reported with Strattera use (Bangs et al., 2008a). Straterra also carries a black-box warning for its capacity to induce suicidal ideation; however, the actual rate of new-onset suicidal ideation or self-harm events reported in the clinical trials was well under 1% (Bangs et al., 2008b). It has a similar cardiovascular profile as stimulants (Vaughan, Fegert, & Kratochvil, 2009), so monitoring of blood pressure and pulse is recommended, as is prescreening for structural heart defects and a family history of sudden cardiac death.

Alpha Agonists

Alpha agonists, including clonidine and guanfacine, comprise the other main class of nonstimulant ADHD medications. Alpha agonists were originally used to treat high blood pressure, but for the past 20 years they have been prescribed off-label for ADHD in children (Connor, Fletcher, & Swanson, 1999; Scahill, 2009). Extended-release versions of clonidine (trade name: Kapvay) and guanfacine (trade name: Intuniv) are now approved by the FDA for the treatment of pediatric ADHD. These developments have been associated with resurgence in the use of alpha agonists for ADHD.

Clonidine was the first agent of this class to be widely used for ADHD. However, the oral formulation of clonidine has a very short duration, necessitating up to four times a day dosing, which often led to adherence

problems and uneven effects throughout the day. In addition, clonidine produces sedation that is severe enough that it is sometimes prescribed as an off-label sleep induction aid in ADHD (Prince, Wilens, Biederman, Spencer, & Wozniak, 1996). Extended-r clonidine produced a significantly greater reduction in inattentive and hyperactive symptoms than placebo, with an estimated standardized mean difference effect size of 0.70 (Jain, Segal, Kollins, & Khayrallah, 2011). It can be dosed once a day and may produce improved tolerability compared with the short-acting version (Jain et al., 2008).

Guanfacine (trade name: Tenex) is an alpha agonist that can be less sedating than clonidine and has a longer duration that allows for twice daily dosing. Extended-release guanfacine (GXR) is approved for treatment of ADHD in patients ages 6 to 17, with some evidence of greater effects in children than in adolescents (Biederman et al., 2008). Effects have been found for 12 hours after morning dosing, and follow-up studies show therapeutic effects maintained over 24 months (Sallee, McBurnett, & Wigal, 2010).

Interest in combining stimulants with alpha agonists has been especially high because they possess very different mechanisms of action and largely opposing side-effect profiles (Childress, 2012). Large double-blind industry-funded trials have been conducted to test whether response to stimulant medication can be improved by the addition of extended-release clonidine or guanfacine. In these studies, adding an alpha agonist to a CNS stimulant improved response over stimulants plus placebo by 25 to 33% without worsening side effects (Kollins et al., 2011; Wilens et al., 2012), leading the FDA to approve their use with CNS stimulants for the combination treatment of ADHD. However, polypharmacy in children should always be considered as a second-line treatment approach used primarily for individuals who are not well controlled after multiple trials of individual medications, behavior therapy, or the combination of these treatments. In addition, tolerability should always be carefully assessed when combining ADHD medications. In contrast to the promising results of combining stimulants with alpha agonists, the combination of stimulant medication with atomoxetine has not been associated with improved outcomes and appears to worsen tolerability (Carlson et al., 2007; Hammerness et al., 2009; Wilens et al., 2009).

The most common side effects associated with alpha agonists are sedation and headaches (Wilens et al., 2012). In one study, 8% of subjects dropped out of treatment due to fatigue (Biederman et al., 2008). Sedation seems to be more common at higher doses but may diminish after the first three weeks of treatment. There are also several rare but serious risks associated with this class of medication, including hypotension, bradycardia, and syncope. In one study, bradycardia or hypotension was reported in 5% of patients, and syncopal events were reported in 2% of patients (Muir & Perry, 2010).

Behavior Therapy

The difficulties associated with ADHD are typically observed in multiple settings, such as at home and at school and with the peer group. As a result, behavioral treatments (BTs) for ADHD have been studied in all of these settings, and the EST Task Force reviewed this research separately (Pelham & Fabiano, 2008; Pelham et al., 1998). Results showed that BT is effective for children with ADHD in home settings by means of behavioral parent training and in school settings by means of classroom contingency management. In parent training programs, parents of children with ADHD are taught principles of behavior management in a series of 8 to 20 meetings delivered in individual or group settings. Parents learn to put the behavioral principles into practice by implementing techniques such as time out, point systems, and contingent attention. In classroom contingency management programs, therapists use a series of consultation sessions at the child's school to develop and implement behavioral techniques that teachers and other school personnel implement. These techniques may include developing specific and concrete rules for the child to follow, developing a system for tracking rule adherence and rule violations, and developing positive and negative consequences that are contingent on the child's performance during school. BT delivered in recreational settings, such as in a Summer Treatment Program, also met the Task Force criteria as an EST. BT is typically effective in reducing the core symptoms of ADHD, as well as related problems such as academic productivity and classroom behavior. This seems to be true regardless of the specific form of BT used. Effect sizes show that changes produced by BT are generally moderate to large, with the largest effect sizes found in studies that used within-subjects designs (Fabiano et al., 2009).

There are several limitations of BT for ADHD. First, although BT frequently improves the behavioral functioning of children with ADHD, treatment effects are rarely sufficient to "normalize" children. Second, treatment gains are often limited to the period during which treatment is actively implemented. Third, not all children with ADHD respond positively to BT. Often, failure to respond is attributable to factors external to the child, such as parents' and teachers' inability or unwillingness to implement the treatment program or therapists' lack of knowledge or skills. In addition, the most effective BT programs are often the most comprehensive and intensive and therefore the most difficult to implement consistently and continually. These factors suggest that, even though BT is not effective for all children with ADHD, some of its ineffectiveness can be explained by the fact that nonresponding children do not actually receive the treatment as intended, do not receive the treatment consistently, or do not receive an adequate dose of the treatment. Such situations are akin to receiving medication at nontherapeutic dosage levels or receiving medication inconsistently. Finally, there is no evidence that BT produces long-term changes

in the negative outcomes often associated with ADHD (Pelham & Fabiano, 2008).

Combined Treatments

The combination of BT and stimulant medication is also empirically supported (Pelham & Fabiano, 2008). A number of arguments can be made in favor of combining stimulant medication and behavioral approaches to treat ADHD. First, combined behavioral and stimulant treatments may show complementary effects, with each addressing the weaknesses of the other. For example, behavioral treatments can be used 24 hours a day, whereas medication treatments may not be used for at least part of the day (e.g., evening and weekends). Similarly, medication treatments may address problems that occur in the absence of adult authority figures or that occur at a low rate, whereas behavioral treatments may be ineffective with these difficulties (Hinshaw, Heller, & McHale, 1992; Hinshaw, Henker, Whalen, Erhardt, & Dunnington, 1989). Second, when behavioral and medication interventions are implemented together, they may amplify each other. Some evidence supports this possibility by showing that behavioral interventions are more effective when combined with medication and that less medication is needed to show positive effects on children's behavior when combined with behavioral interventions (Pelham, Burrows-MacLean, Gnagy, Fabiano, Coles, et al., 2014; Pelham et al., 1993). Third, combined treatments may be more cost-effective because less treatment is needed to produce the same outcome.

Despite this strong rationale, several caveats and limitations should be noted. First, combined stimulant and behavioral interventions are relatively unstudied. Second, the incremental benefits of combined treatments do not appear to persist when either component is discontinued (Pelham et al., 1988). Third, studies of combined treatments have focused on acute effects rather than on long-term maintenance. It is not known whether combined BT and medication produces improved long-term outcomes relative to either treatment alone, although the MTA study (discussed later) has provided some important information about this topic. Finally, previously combined studies had only examined stimulant medications, but they have now begun to examine the effects of BT and nonstimulants as well. For example, the addition of behavior modification to Strattera is associated with improved functional outcomes as well as improved parent satisfaction with treatment (Waxmonsky et al., 2010).

Multimodal Treatment for ADHD Study

We end our discussion of ESTs with a brief review of the Multimodal Treatment for ADHD (MTA) study. This study is relevant because it is the largest trial ever conducted on behavioral and stimulant treatments for ADHD

and because it has arguably had an enormous influence on treatment practices for ADHD. The MTA study was initiated in the early 1990s (Richters, Arnold, Jensen, Abikoff, Conners, et al., 1995) and enrolled 579 boys and girls with ADHD, ages 7.0 to 9.9, who were assessed and treated at seven locations across North America (Hinshaw, March, Abikoff, Arnold, Cantwell, et al., 1997). Children were randomly assigned to one of four treatment conditions: (1) stimulant medication only; (2) behavioral treatment only; (3) both stimulant medication and behavioral treatment; and (4) standard community care. These treatments were provided for 14 months, after which outcomes were evaluated in numerous domains. The main questions of interest were: (1) What are the relative efficacies of behavioral and pharmacological treatments? (2) What is the incremental benefit of combining these treatments over either treatment alone? and (3) How do these evidence-based treatments compare with treatments routinely given in the community? (MTA Cooperative Group, 1999a).

Results of data collected immediately after the end of treatment were widely interpreted as showing that stimulant medication treatments produced superior outcomes to behavioral treatments, with little added benefit of behavioral treatment over and above stimulant medication (MTA Cooperative Group, 1999a). Treatment effects were generally not moderated by age, comorbidity, or other factors (MTA Cooperative Group, 1999b). These results were widely disseminated in scientific and popular press outlets (e.g., CNN, *The New York Times*, CBS), often boiled down to a simple take-home message such as "The study found unequivocally that medication . . . was significantly more effective at treating the ADHD symptoms than treatments which did not include medication" (PBS *Frontline*, 2001). Nevertheless, not all data were consistent with this conclusion (Pelham, 1999a, 1999b, 2000). For example, although treatments that included medication produced superior improvement on measures of ADHD symptoms, treatments that included behavioral techniques produced equivalent or better improvement on measures of impairment, such as observations of classroom behavior, parent and teacher ratings of social skills, parent ratings of parent–child relationships, peer sociometric ratings, and academic achievement measures.

Importantly, assessments of treatment effects conducted since the initial results have shown a consistent pattern of diminishing differences between the treatment groups. As noted in the most recent publication from this study, which examined outcomes 8 years posttreatment:

> In nearly every analysis, the originally randomized treatment groups did not differ significantly on repeated measures or newly analyzed variables (e.g., grades earned in school, arrests, psychiatric hospitalizations, other clinically significant outcomes). . . . The MTA participants fared worse than the local normative comparison group on 91% of the variables tested. (Molina et al., 2009, p. xx)

These results provide a sobering reality check; empirically supported treatments for ADHD may be the best of current options and produce significant and meaningful acute improvements, but they are also apparently not sufficient for ameliorating negative long-term outcomes. This conclusion highlights the need to seek other potential treatments for ADHD. We turn our attention to these next.

PROMISING TREATMENTS

Recent years have witnessed dozens of efforts to develop and evaluate new approaches for treating ADHD, due in part to the limitations of established treatments and to public interest in and use of alternative treatments (Bussing, Zima, Gary, & Garvan, 2002). Research on some of these new treatment approaches has produced promising results. For purposes of the present discussion, we consider promising treatments to be those that: (1) have been subjected to methodologically sound research that (2) reports positive treatment effects in at least some of these studies but (3) do not currently meet full criteria as ESTs. We are not advocating or recommending any of these as primary treatments for ADHD at this time; inclusion in this category should not be taken as advocating for the efficacy of any of these treatments. Instead, inclusion in this category simply means that the jury is still out regarding whether the treatment has potential benefits, if potential benefits outweigh costs, and how these treatments compare with established treatments. More and better research is needed on all of these approaches before they can be either recommended or ruled out.

Peer-Directed Interventions

Most classroom-based treatments for ADHD are teacher-directed. However, peers have been recruited to assume roles as behavior change agents by monitoring and reinforcing desirable social and academic behaviors of ADHD students, prompting appropriate behavior, or serving as tutors for remediating specific academic skills. DuPaul and Stoner (1994) argued that peer tutor interventions target important skills typically unaffected by traditional contingency management programs and satisfy the learning needs of many ADHD students by providing frequent, immediate feedback on important aspects of educational functioning. The few studies evaluating this intervention approach have used case-control methods. These studies find that peer tutoring results in significant improvements in both classroom behavior and academic performance in elementary school students with ADHD (DuPaul, Ervin, Hook, & McGoey, 1998; DuPaul & Henningson, 1993), as well as improved social functioning (Plumer & Stoner, 2005).

Self-Directed Interventions

Self-monitoring and self-reinforcement of behavioral interventions involve having children monitor and evaluate their own social or academic performance according to prescribed criteria, and rewarding themselves based on those self-evaluations. These techniques have been used with ADHD students to increase on-task behavior during individual seatwork (Barkley, Copeland, & Sivage, 1980), improve reading comprehension (Edwards, Salant, Howard, Brougher, & McLaughlin, 1995), and increase cooperative play interactions (Hinshaw, Henker, & Whalen, 1984). A meta-analytic review of this research identified 16 studies that collectively enrolled 51 participants, the majority of whom were elementary school-age children (Reid, Trout, & Schartz, 2005). The average effect sizes from these studies showed that self-direct interventions produced large increases in on-task behavior (standardized mean difference effect size = 1.6), and academic productivity (standardized mean difference effect size = 1.1) and large decreases in inappropriate behavior (standardized mean difference effect size = 1.3). These promising results are tempered by the fact that nearly all studies reviewed used single-case experimental designs; there is as yet no randomized clinical trial.

Neurofeedback

Electroencephalogram (EEG) studies of children with ADHD have shown that they often display an excess of slow theta waves and lower than expected levels of beta waves. Further, treatment with CNS stimulants leads to increased beta and decreased theta waves (Lofthouse, Arnold, Hersch, Hurt, & DeBeus, 2012). Based on these and other findings, it has long been hypothesized that EEG neurofeedback may correct the imbalance of beta to theta waves, resulting in improvements in the brain's capacity for self-regulation. Neurofeedback treatments use video, audio, or even tactile sensory feedback in an attempt to reduce hyperactivity, inattention, and impulsivity by training children with ADHD to increase brain wave activity associated with sustained attention and decrease brain wave activity associated with inattention and daydreaming.

Neurofeedback has been touted as a treatment for ADHD for over 30 years (Lubar & Shouse, 1977). Proponents have often made strong claims for its effectiveness as a treatment for ADHD. For example, in one article published in *Psychology Today* (Robins, 1998), EEG treatments for ADHD (as well as EEG treatments for epilepsy, closed head injury, chronic substance abuse, and posttraumatic stress disorder) were described as effective and supported by research. The author backed these claims by providing case study examples of how EEG has been successfully used to treat ADHD. Of note is the author's report that EEG neurofeedback was being used as the primary treatment for ADHD by 22 schools in New

York. "So far," the author reported, "neurofeedback [in these schools] has kept twenty students out of expensive special-education classrooms and thereby saved the district an estimated $500,000." The author went on to describe his own positive experiences with EEG. Early studies of EEG effects on ADHD seemed to support this enthusiasm but were difficult to interpret because they used open trial (uncontrolled) methods. Even so, several recent randomized controlled trials (RCTs) have reported medium-sized effects compared with various control conditions (e.g., cognitive self-control therapy, biofeedback focused on relaxing forehead muscles) across a variety of measures (parent-rated symptom scales, neuropsychological batteries), with larger effects for measures of impulsivity and inattention than for measures of hyperactivity (see Arns, de Ridder, Strehl, Breteler, & Coenen, 2009, for a review).

Despite these seemingly positive results, there is reason for caution (Kline, Brann, & Loney, 2002; Lohr, Meunier, Parker, & Kline, 2001). Typically, studies that report positive effects of EEG biofeedback as a treatment for ADHD are characterized by methodological problems, such as confounded treatments (Boyd & Campbell, 1998), inconsistent use of dependent measures among subjects (Rossiter & LaVaque, 1995), absence of clinically meaningful dependent measures (Wadhwani, Radvanski, & Carmody, 1998), and nonstandardized collection of posttreatment measures administered anywhere from 0 to 12 months following cessation of treatment (Alhambra, Fowler, & Alhambra, 1995). Further, many of these studies demonstrate changes in *laboratory* measures of ADHD, but there is no evidence that these changes correspond to changes in real-life settings.

Perhaps the largest concern is that the studies rarely use appropriate blinding of the experimenter, child, and/or informant (parent, teacher), and may thus be capitalizing on placebo effects and other artifacts, which can be significant problems in ADHD treatment research (Waschbusch, Pelham, Waxmonsky, & Johnston, 2009). The importance of blinding is demonstrated by a recent review that found that neurofeedback studies that incorporated more rigorous blinding procedures produced smaller, often nonsignificant treatment effects (Lofthouse et al., 2012). As one example, a recent study examined EEG effects using a videogame paradigm (Arnold et al., 2012). Participants were children with ADHD (n = 39), all of whom played a videogame for 40 treatment session. The key manipulation was that the videogame for children in the treatment condition was contingent on their EEG activity (i.e., it incorporated neurofeedback), whereas the videogame for children in the control condition did not. Results showed that (1) the control condition effectively blinded the participants to treatment assignment, with neither parents nor children able to identify the assigned condition (treatment vs. control) at rates better than chance; and (2) neurofeedback performed no better than the sham control across many outcome measures and performed nominally worse

on some outcome measures. When combined with the fact that this treatment can be labor-intensive and costly (with few insurance companies currently covering it), these results raise important questions about the utility of neurofeedback for ADHD.

Furthermore, not all children with ADHD have an excess of theta waves; this suggests that neurofeedback may only be efficacious for a subset of ADHD youth. Even so, the fact that other recent randomized trials that control for placebo effects have reported modest positive effects (e.g., Gevensleben et al., 2009) indicates that the jury is still out on neurofeedback as a treatment for ADHD.

Cognitive Treatments

Cognitive treatment for children with ADHD was first proposed over 40 years ago (Meichenbaum & Goodman, 1971). Many types of cognitive treatments (also called cognitive-behavioral treatments) have been applied to children with ADHD, including training in verbal self-instruction, problem-solving strategies, and cognitive modeling (Abikoff, 1987, 1991; Braswell & Bloomquist, 1991). The rationale for these treatments is that behavioral self-control can be increased by enhancing specific cognitive or metacognitive skills that are believed to underlie and promote impulse control, goal-directed behavior, or both. Because the seeming absence or inefficiency of such internal mediators appears to characterize children with ADHD, cognitive treatments enjoy a strong intuitive appeal.

Nevertheless, numerous studies have found that cognitive treatments are ineffective for ADHD. For example, one research team administered 16 weeks of intensive cognitive training to children with ADHD and found no differences on multiple academic, cognitive, and behavioral measures in comparison with attention control (i.e., providing support but no active intervention) and no-training groups (Abikoff et al., 1988). Other studies report similar negative findings (Bloomquist, August, Cohen, Doyle, & Everhart, 1997; Brown, Wynne, & Medenis, 1985), even though there has been demonstrated efficacy of cognitive training programs for other childhood disorders (Dujovne, Barnard, & Rapoff, 1995; Kendall & Gosch, 1994; Lochman, 1992).

Despite these disappointing results, we include cognitive interventions as promising because there may be several exceptional circumstances under which these programs have some clinical usefulness. First, cognitive training that focuses on improving social skills and that is adjunctive to operant behavioral or clinical behavioral interventions may be beneficial (Pelham et al., 1988). Second, anger control training in the context of intensive behavioral interventions may also be useful (Hinshaw, Buhrmester, & Heller, 1989; Hinshaw et al., 1984; Lochman, Boxmeyer, Powell, Barry, & Pardini, 2010). Third, problem-solving training may be helpful for children with ADHD who have comorbid aggression, especially if cognitive training

is combined with parent training (Kazdin, Bass, Siegal, & Thomas, 1989; Kazdin, Esveldt-Dawson, French, & Unis, 1987; Lochman, 1992).

We also include cognitive therapies in the promising category because of more recent forms of cognitive treatment that have focused on improving executive functioning skills, such as working memory and self-control, which are clearly associated with ADHD. These treatments use several methods, including computer-based programs such as CogMed (Klingberg et al., 2005), physical activity (Best, 2012), and school-based curricula (Riggs, Greenberg, Kusché, & Pentz, 2006). There is early evidence from RCTs that each of these approaches can improve aspects of executive functioning (Diamond, 2012). For example, a recent RCT of adolescents with ADHD reported that those who received the CogMed intervention had significantly improved working memory abilities compared with those who completed a control task (Gray et al., 2012). Studies of other computer-based executive functioning programs have reported similar results (Rabiner, Murray, Skinner, & Malone, 2010).

At the same time, these results, though promising, must be interpreted in light of several important caveats. First, studies that report positive treatment effects are offset by similarly designed studies that do not find treatment effects (see Chacko et al., 2013, for a review of CogMed studies). Second, the positive effects induced by cognitive training programs typically do not generalize beyond the specific skills targeted by the programs (Diamond, 2012; Shipstead, Hicks, & Engle, 2012; Silberg et al., 1996). Third, effects that do emerge may not be stable over time. For example, one meta-analysis found that interventions aimed at improving working memory significantly improved both verbal and visuospatial working memory in the short term (standardized mean difference effect sizes were 0.79 and 0.52, respectively), but effects on the same measures were small or nonsignificant a few months later (Melby-Lervag & Hulme, 2012). This same review also reported that working memory interventions improved only the specific working memory skills targeted and did not benefit other aspects of working memory. Working memory training programs have been criticized on other methodological grounds as well (Shipstead, Redick, & Engle, 2012).

More directly relevant, evidence that executive functioning training programs are effective for ADHD is, at best, inconsistent (Chacko et al., 2013). For instance, one study of CogMed reported a reduction in hyperactive symptoms (Klingberg, Forssberg, & Westerberg, 2002), but this effect failed to replicate in a later study (Klingberg et al., 2005). Similarly, a recent study reported that working memory training significantly improved symptoms of ADHD measured using observations during an academic task completed in an experimental situation, but did not improve ADHD symptoms measured using parent report (Green et al., 2012). Thus, executive functioning training programs are promising and warrant further evaluation but cannot be recommended at this time.

Dietary Restriction of Artificial Food Colorings and Preservatives

Perhaps more than any other individual, Feingold (1973, 1975a, 1975b, 1976) popularized the notion that ADHD symptoms result from a toxic reaction to certain low-molecular-weight chemicals (called salycilates) ingested through the diet. Salycilates occur naturally in some foods and artificially in other foods via colorings and flavorings. Feingold suggested that elimination of these substances through dietary restriction would improve the behavioral and cognitive functioning of up to 60% of children with ADHD (Feingold, 1975a, 1975b). Although Feingold's writings never fully delineated the mechanisms by which toxic reactions occurred, his early statements were emphatic, generating much debate and a flurry of research activity that continued for many years.

The first controlled investigation of dietary management used a counterbalanced design to compare the Feingold diet (i.e., a diet that excludes offending substances specified by Feingold) with a control exclusionary diet (Conners, Goyette, Southwick, Lees, & Andrulonis, 1976). Results were mixed, showing a modicum of support for the Feingold Diet and suggesting that continued investigation of effects of the diet on children with ADHD was warranted. These mixed results foreshadowed findings from experimental studies conducted over the next several years. On measures of ADHD symptoms, some studies reported positive effects of the Feingold Diet (e.g., Cook & Woodhill, 1976; Holborow, Elkins, & Berry, 1981; Rapp. 1978), whereas other studies reported either minimal or no effects (e.g., Conners, 1980; Mattes & Gittleman-Klein, 1978; Stine, 1976). These initial studies were characterized by a number of methodological problems. For example, improvements in ADHD symptoms were typically detected only on parent ratings (many of whom were not blind to the treatment conditions), suggesting they may have been attributable to placebo effects or other artifacts (Baker, 1980). Research that accounted for these methodological flaws did not find clinically meaningful gains (Harley & Matthews, 1978; Kavale & Forness, 1983; Mattes, 1983), and subsequent rebuttals were generally unconvincing (Rimland, 1983; Weiss, 1982).

More recent studies have continued to test the effects of artificial colorings (ACs) and preservatives, employing more stringent experimental procedures. The results of some of these studies have been more promising (Rowe, 1988; Rowe & Rowe, 1994). For example, Bateman and colleagues (2004) recruited 387 subjects from the community and divided them into four cohorts based on the presence of ADHD as well as atopy (allergies). For the first week of the study, parents of all participants gave their child a diet free of ACs and benzoates. Over the next three weeks, participants were randomly assigned to one week of placebo, followed by a one-week washout and then one week of a drink containing 20 mg per day of benzoate and ACs, or the reverse order. Treatment effects were measured using parent ratings, as well as direct observations of attention, activity level, and

impulse control in a laboratory setting. Results of parent ratings showed that hyperactivity improved in subjects with and without ADHD when ACs and benzoates were removed, worsened when the ACs were reintroduced, and improved again upon subsequent removal, yielding a moderate effect size of 0.51. The presence of atopy at baseline did not influence these findings, suggesting that the impact of ACs and benzoates on hyperactivity is not mediated through allergic pathways. However, these results were not replicated in the laboratory observations, raising questions about the robustness of the effect and the role of parent expectancy biases (even though a placebo condition was included).

McCann and colleagues (2007) sought to replicate and extend the Bateman study by recruiting both preschoolers and elementary school-age participants and assessing two doses of ACs. Participants were recruited from the general population; neither ADHD nor atopy was required for enrollment in the study. The first dose of ACs was identical to that used in the Bateman study, whereas the second was created to match the average daily consumption of ACs, preservatives and flavorings in the diet of modern British children, with doses of both adjusted for participant weight. The assessment measures were also expanded relative to the Bateman study. Results showed that exposure to ACs and sodium benzoate was associated with a small but significant worsening on a global hyperactivity index (a combination of parent, teacher, and clinician ratings), yielding standardized mean difference effect sizes of 0.20 or less for both preschool and elementary-age children. Largely as a result of these and other studies, Europe has recently required warning labels for foods that use some artificial colorings and has requested voluntary bans on others (Stevens, Kuczek, Burgess, Hurt, & Arnold, 2011).

A recent meta-analysis quantitatively summarized several studies and drew conclusions that were generally consistent with this research (Nigg, Lewis, Edinger, & Falk, 2012), but suggesting that several important caveats must be kept in mind when interpreting these studies. First, the effects of dietary interventions, even when statistically significant, are typically small in magnitude. Specifically, the five most methodologically rigorous studies that examined dietary interventions for ADHD symptoms collectively produced a standardized mean effect size of 0.29. As Barkley (2012, p. 2) noted in reviewing this research, "Though significant, these effects are rather small . . . the meaning of these results as a basis for advising a restrictive diet to manage ADHD is open to some question." Second, the meta-analysis estimated that although 33% of children seem to respond positively (to some degree) to dietary interventions, only 8% of children with ADHD appear to have symptoms related to food color effects. Why more children respond to restricting ACs than show a negative reaction to ACs remains unclear. Third, the number of well-conducted studies is small. In the Nigg et al. (2012) meta-analysis, only five studies that cumulatively enrolled 164 participants contributed data to the dietary restriction effect

size. Fourth, there was significant and meaningful variability, both within and across studies, when examining the effects of food colors/additives on children's behavior. Specifically, the review found evidence for significant but small negative effects of ACs when behavior was measured by parent or lab tasks but not when behavior was measured by teachers or observers. Likewise, there were no significant effects when studies were restricted to those that used colors/additives approved by the FDA. Overall, these studies suggest that both tentative optimism and caution are warranted when interpreting these studies, and so they argue for further research.

Nutritional and Dietary Supplements

The list of proposed nutritional supplements for treatment of ADHD is large. Seemingly every supplement that has come to market has been suggested as a possible treatment of ADHD, such as amino acids, vitamins, zinc, magnesium, and phytochemicals (Curtis & Patel, 2008). In general, consumers should be wary of efficacy claims as the FDA regulations for nutritional supplements are much less stringent than those for pharmaceuticals. Nor should consumers assume that supplements are safe simply because they are "natural."

One supplement that may show promise, however, is omega 3 fatty acids. Omega 3 fatty acids are associated with cell membrane stability and must be consumed in the diet because humans do not synthesize them (unlike omega 6 fatty acids). A recent meta-analysis of omega 3 studies for ADHD found 10 trials that collectively enrolled 699 subjects (Bloch & Qawasmi, 2011). Overall, omega 3 treatment was associated with a small reduction for clinician-rated ADHD symptoms (standard mean difference effect size of 0.31), with a similar reduction for parent-rated symptoms (0.29). These effects are about half the size of treatment effects typically produced by nonstimulant medications and one-third to one-quarter the effect typically produced by stimulant medications (Greenhill & Ford, 2002). Most studies had important design limitations, however, such as not accounting for attrition and failing to collect teacher ratings. Nevertheless, almost no side effects of omega 3 emerged. Given the modest potential benefit and the low apparent risk profile, further investigation of this treatment seems warranted.

UNSUPPORTED TREATMENTS

We define unsupported treatments as those that have been examined in research and should not be recommended for the treatment of pediatric ADHD due to lack of positive effects, unacceptably high risks, or both. We review several hypothesized treatments for ADHD that have been evaluated with at least some research and shown to be ineffective or otherwise

unacceptable. Our review of unsupported treatments excludes treatments not subjected to methodologically sound research, such as eye movement desensitization and reprocessing (Tinker & Wilson, 1999), homeopathic treatments (Carper, 1998; Reichenberg-Ullman & Ullman, 1996), stability balls (Fedewa & Erwin, 2011), dolphin-assisted therapy (Marino & Lilienfeld, 2007), gum chewing (Pham, 2011), meditation (Krisanaprakornkit, Ngamjarus, Witoonchart, & Piyavhatkul, 2010), weight vests (Olson & Moulton, 2004), and many others. As this selective list suggests, there are simply too many such "treatments" to feasibly review in full. In general, these purported interventions have either not been studied or have been studied using methodologically flawed designs characterized by small sample sizes, insufficient control groups, lack of randomization, insufficient subject and experimenter blinding, and absent or inappropriate control groups. As a consequence, we focus on the most widely studied unsupported interventions.

Antidepressants

Several investigations have examined antidepressants, most notably bupropion and the tricyclic antidepressants, as treatments for ADHD. A recent review of antidepressants for adults with ADHD concluded that there is limited evidence that they are effective, but that their effects are substantially smaller than those of stimulant treatments (Verbeeck, Tuinier, & Bekkering, 2009). Although tricyclic antidepressants have been found to produce positive effects for the treatment of ADHD in children (Biederman, Baldessarini, Wright, Knee, & Harmatz, 1989; Biederman, Gastfriend, & Jellinek, 1986), all antidepressants (1) carry a black-box warning for the inducement of suicidal ideation in children and adolescents (Birmaher & Brent, 2007), (2) have a slower therapeutic onset, and (3) require daily dosing. Taking all these factors into consideration, there is insufficient justification for using antidepressants for ADHD, given that multiple other options with better safety and efficacy profiles exist.

Dietary Restriction of Sugar and Sweeteners

Refined sugar has long been proposed as a primary cause of hyperactivity and other child behavior problems (Smith, 1975), and a sizable portion of parents continue to endorse this view (Gilmore, 2010). However, well-controlled studies have not demonstrated an effect of sugar on children's behavior, even among children thought to be sugar-sensitive (Milich, Wolraich, & Lindgren, 1986). Similar conclusions have been reported in controlled studies of aspartame on behavior (Wolraich, Wilson, & White, 1995). For example, Wolraich and colleagues (1988, 1994) compared three diets that had various levels of sucrose and sweetener (high sucrose–low sweetener, low sucrose–high sweetener, and placebo) in two groups of

children presumed to be especially vulnerable to the effects of sugar ingestion (i.e., preschool and school-aged children nominated by parents as having highly adverse behavioral reactions to sugar). The diets were presented in 3-week blocks using a counterbalanced, double-blind, crossover design. Results showed no differences among the three diets on any of almost 40 behavioral and cognitive measures. Further, there is evidence that the only significant influence of sugar on children's behavior is a nocebo (reverse placebo) effect; that is, parents who believe their child has just ingested sugar rate their child as more behaviorally disruptive and are simultaneously more controlling and negative toward their child compared with parents of children who believe their child did not ingest sugar. These effects are independent of whether or not the child actually received sugar (Hoover & Milich, 1994). Collectively, this evidence demonstrates that neither sugar nor aspartame ingestion exerts appreciable effects on children's behavior and that restricting them is not likely to have an appreciable effect on symptoms of ADHD, although it may have significant effects on other health outcomes, such as obesity and diabetes (Lustig, 2012).

Sensory Integration Interventions

Occupational therapist A. Jean Ayers (1979) proposed that an important task of the human neurological system is to organize and integrate sensory information from the environment and that deficits can arise as this process develops in children. Based on this assertion, she and others developed sensory integration interventions (SIIs), which have become some of the most common interventions occupational therapists use when working with children (Shaaf & Davies, 2010). SIIs involve activities that are believed to help children organize their sensory system by providing vestibular, proprioceptive, auditory, and tactile inputs, typically using tools such as brushes, swings, balls, and other specially designed equipment (American Association of Pediatrics, 2012). Occupational therapists commonly use SII as an intervention for Autism (Lane & Schaaf, 2010), but its application to ADHD is also common (Koenig & Rudney, 2010). Indeed, estimates suggest that 90% of occupational therapists who work in school settings use SII to address learning disabilities, ADHD, autism, and behavior problems more generally (May-Bensen & Koomar, 2010). It is not difficult to understand why SII has been applied to treat ADHD, because impulsivity and distractibility are considered two key symptoms of sensory integration problems. For example, in describing the typical school problems experienced by a child with sensory integration dysfunction, Ayers and Robbins (2005, p. 11) wrote that "in the classroom, he may be easily distracted by all the sounds, lights, and confusion of people doing different things. His brain is overly stimulated and it responds with a lot of excessive activity. The hyperactive child 'jumps all over the classroom,' not because that is what he wants to do but because his brain is running out of control. The

confusion in his brain makes it impossible to focus or concentrate, and so he can't understand what his teacher is teaching."

Despite the widespread application of SII to treat ADHD, the evidence to support this approach is lacking. As cogently summarized in a recent policy statement by the American Academy of Pediatrics (2012), there is as yet no good evidence to support the use of SSI for children with ADHD or other behavior disorders. Further, recent single-case research suggests that SSI may actually impede more effective interventions—namely, behavioral interventions, although this study was conducted on children with autism rather than ADHD (Devlin, Healy, Leader, & Hughes, 2011).

Play Therapy

Play therapy originated with work by Anna Freud and typically consists of therapists encouraging children to act out feelings, thoughts, and experiences in the context of play activities, usually in one-on-one therapy sessions (Bratton & Ray, 2000). Play therapy is widely used in school and mental health settings (Ray, Armstrong, Warren, & Balkin, 2005), and it has been suggested to be an effective treatment for children with ADHD (Kaduson, 1997). Although play therapy may be effective for some child problems (Bratton & Ray, 2000; Bratton, Ray, Rhine, & Jones, 2005), it does not appear to be effective for the treatment of ADHD. One randomized trial has examined the use of play therapy for the treatment of ADHD and found no difference on symptoms of ADHD at the end of treatment relative to a control condition (Ray, Schottelkorb, & Tsai, 2007). Thus, there is no evidence to support the use of traditional play therapy for the treatment of ADHD. It should be noted, however, that nontraditional forms of play therapy may produce some benefit. As noted by O'Neill, Rajendran, and Halperin (2012), animal studies suggest that rough-and-tumble play during development may have beneficial effects on reducing inattention and overactivity (Panksepp, 1998; Panksepp, Burgdorf, Turner, & Gordon, 2003). In addition, as reviewed earlier, emerging evidence suggests that certain games that exercise children's executive functioning abilities (such as "Simon says") may yield benefits when delivered in critical developmental stages (Diamond, 2012; Diamond, Barnett, Thomas, & Munro, 2007). Likewise, play may have adjunctive benefits in the treatment of ADHD by providing a child-friendly means of delivering behavioral treatments in the context of sporting activities (O'Connor, Fabiano, Waschbusch, Belin, Gnagy, et al., 2013) or using playtime to improve the parent–child relationship (O'Neill et al., 2012).

CONCLUSIONS

ADHD is a chronic condition characterized by hyperactivity, impulsivity, and inattention that is associated with serious impairment in interpersonal,

academic, and behavioral performance. As it has for the past three decades, evidence continues to suggest that stimulant medication, classroom behavior interventions, and parent behavior management training should be considered the first-line treatments for ADHD. Other treatments for ADHD also have empirical support (e.g., nonstimulant medications), but their effects are not as large as those of first-line treatments. Many other treatments have been proposed, but because of the extensive body of knowledge supporting well-established treatments for ADHD, those who propose new treatments bear a "burden of proof" to provide convincing evidence that the cost–benefit ratio of the proposed treatments equal or outweigh those of empirically supported treatments. Given that many treatment options for ADHD have empirical support, the onus is on proponents of new treatments to demonstrate that they are effective, have an acceptable risk profile, and exhibit advantages over established ESTs.

Curiosity and interest in alternative treatments remain high (Bussing et al., 2002), despite a dearth of evidence for their efficacy and a lack of information about their risk profiles. There are a number of reasons for this state of affairs. First, some parents and professionals hold negative attitudes and beliefs regarding the desirability of established treatments even before trying them. A priori negative attitudes seem to be especially widespread for medication treatment (Waschbusch et al., 2011). Such negative attitudes toward stimulant medication may be more common among parents than among professionals (such as classroom teachers). For example, in the MTA study, parents gave higher consumer satisfaction ratings to treatments that included behavioral components, whereas this trend was not found for teachers (Pelham, Gnagy, Greiner, & MTA Cooperative Group, 2000). These different attitudes toward treatment could lead to considerable tension, potentially undermining the effectiveness of any treatment that is administered. A more fruitful approach would be for both parents and teachers to suspend their a priori beliefs about ESTs for ADHD and allow the child's response to guide the determination about treatment.

Second, even when such attitudes or beliefs are absent prior to treatment, the limitations of established treatments (discussed earlier) often arise once treatments are implemented. Side effects of stimulants are typically mild and manageable, but can dampen enthusiasm for this form of treatment. Likewise, behavioral treatments require considerable effort and organization to be effective. These factors can lead parents to become disenchanted with ESTs, thereby spurring interest in alternative treatments.

Third, there are different methods of delivering information about ESTs compared with promising or unsupported treatments (Waschbusch, Fabiano, & Pelham, 2012). Proponents of ESTs tend to be scientifically oriented professionals who are (to a greater or lesser extent) trained to be skeptical and cautious in their claims of treatment effectiveness. Advocates of other treatments often do not have such constraints and may be financially

or otherwise motivated to make exaggerated claims of treatment effects. For example, one day after the shootings at Columbine High School, the Cable News Network (CNN) repeatedly broadcast interviews with an "expert" who claimed that Ritalin was linked to school violence (Seay, 1999). Yet CNN apparently neglected to report that this "expert" had authored a book arguing against the use of Ritalin (she has also authored a book arguing against the use of amoxicillin); the research on which she was basing her claims was anecdotal; and the institute she directed was founded by the Church of Scientology, a well-known opponent of psychiatry and psychiatric medications (Seay, 1999). The result of these different approaches for describing treatments is that parents who have no formal training or knowledge about ADHD may be in the position of selecting among treatments that they perceive as (1) entailing a great deal of work (behavioral treatment), (2) having a significant cost (side effects and monetary costs of stimulants), or (3) interventions that have neither of these limitations and promise to cure ADHD.

One reason consumers are faced with this situation is that there is little or no regulation of nonmedication treatments for mental health problems, whereas the same is not true for most medical procedures, including medications (Weisz, 2000). In the absence of such regulation, it is important that the scientific community, including clinicians who use evidence-based approaches, provide clear guidelines to consumers and to media about what is and is not empirically supported. Those who rely on science-based treatments must advocate for the treatments that work and argue strongly against treatments that do not work or that have not been examined. This is not just a nice social convention; using nonsupported treatments can exert serious negative effects on children. For example, one elementary school child we recently treated was referred to our clinic after his mother discovered that his school had been making him wear a weighted vest every time he got "too hyper." The mother found out about the practice only because the child's sibling was asked by classmates why her brother "had to wear a hyper vest," a question the sibling in turn asked the mother. When the mother asked the child with ADHD about the weighted vest, he broke down in tears and described feeling stigmatized by the school's practice.

Fourth, if available longitudinal studies are accurate in their suggestion that stimulants have little effect on long-term outcomes (Molina et al., 2009), some parents and treatment providers may seek out controversial treatments because they become desperate. Given that ADHD is a chronic disorder, it is not difficult to imagine cases in which stimulants, behavioral modification, or both, are used effectively in childhood but not in adolescence due to such factors as refusal to take stimulants or lack of parental control over the contingencies that motivate the youth. In such cases, alternative treatments could offer parents hope that established treatments cannot.

Almost all treatments for ADHD (empirically supported or not) require considerable investment of resources (time, energy, and/or money). Because these resources are limited, ethical considerations require that they be used judiciously and applied to intervention efforts with known risks and benefits. As collaborators in treatment decision making, professionals have both an opportunity and an obligation to (1) present parents with accurate information regarding treatment options for ADHD, (2) base their decisions and recommendations about treatment options on research evidence (rather than on anecdotal stories or their clinical intuition/experience), and (3) advocate strongly for the treatments that are most likely to benefit the child (especially ESTs). A challenge to professionals in meeting this obligation is staying abreast of new findings regarding established treatments for ADHD, not to mention findings regarding newly explored or promising, but largely untested, treatments. In addition, information regarding alternative treatments for ADHD often lies outside of the academic domain, requiring professionals to divide their attention among various information sources and databases. Such challenges are difficult to meet but are fundamental to the advocacy role required of competent professionals.

A number of key questions concerning treatments for ADHD remain to be addressed in future research. First, although ESTs have considerable research backing, little is known about the mechanisms that account for their effectiveness. What aspects of behavioral treatments and stimulants are effective for what aspects of ADHD? Answering these questions could greatly advance the effectiveness of treatments and lead to important advances in our understanding of ADHD. Second, more research is needed on promising treatments for ADHD. As described by our criteria for categorizing these treatments, these treatments boast at least some research support, but there is insufficient evidence to render a decision concerning their efficacy. Third, there is a need for studies examining whether treatments are moderated by factors such as age, genotype, gender, or comorbidity. Fourth, research is needed on long-term effects of treatment. Current information about the maintenance and generalization of treatment effects in children with ADHD is by-and-large based on methodologically poor studies. Further investigation of the long-term outcome of various types and combinations of treatment for ADHD is therefore sorely needed. Fifth and finally, there is a need for research on combined interventions. Although the combination of behavior therapy and stimulant medication has received considerable study, almost nothing is known about other treatment combinations, such as the combination of behavior therapy and nonstimulant treatments. Even within the domain of behavior therapy and stimulant medication, there are important unanswered questions, such as how to sequence the two treatments and what doses of each are optimal when used in combination. Research that provides answers to these questions would greatly benefit children with ADHD.

GLOSSARY

Attention-deficit/hyperactivity disorder (ADHD): A disorder characterized by developmentally inappropriate levels of inattention, impulsivity, and/or hyperactivity. Also referred to as attention-deficit disorder (ADD).

Behavior therapy: A treatment approach that uses environmental factors and learning principles as mechanisms of change.

Multisite treatment of ADHD (MTA) study: A large-scale research study funded by the National Institute of Mental Health to compare the efficacy of stimulant medication, behavior therapy, and the combination of these two treatments for ADHD.

Neurofeedback: A treatment approach that allows individuals to monitor their own brain waves in an effort to alter their behaviors.

Randomized controlled trial (RCT): An experimental procedure in which participants are assigned to a treatment condition or a comparison (usually nontreatment) condition using chance to control for observed and unobserved differences between participants.

Stimulant medication: Drugs that act on the central nervous system, and in particular the frontal–striatal region of the brain, by influencing neurotransmitters; commonly used to treat ADHD.

REFERENCES

Abikoff, H. B. (1987). An evaluation of cognitive behavior therapy for hyperactive children. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 10, pp. 171–216). New York: Plenum Press.

Abikoff, H. B. (1991). Cognitive training in ADHD children: Less to it than meets the eye. *Journal of Learning Disabilities, 24*(4), 205–209.

Abikoff, H. B., Ganeles, D., Reiter, G., Blum, C., Foley, C., & Klein, R. G. (1988). Cognitive training in academically deficient boys receiving stimulant medication. *Journal of Abnormal Child Psychology, 16*, 411–432.

Alhambra, M. A., Fowler, T. P., & Alhambra, A. A. (1995). EEG biofeedback: A new treatment option for ADD/ADHD. *Journal of Neurotherapy, 1*, 39–43.

American Academy of Pediatrics. (2011). ADHD: Clinical practice guidelines for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics, 128*(5), 1–18.

American Association of Pediatrics. (2012). Policy statement on sensory integration therapies for children with developmental and behavioral disorders. *Pediatrics, 129*, 1186–1189.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Arnold, L. E., Lofthouse, N., Hersch, S., Pan, X., Hurt, E., Bates, B., et al. (2013). EEG neurofeedback for ADHD: Double-blind sham-controlled randomized pilot feasibility trial. *Journal of Attention Disorders, 175*(5), 410–419.

Arns, M., de Ridder, S., Strehl, U., Breteler, M., & Coenen, A. (2009). Efficacy of neurofeedback for treatment in ADHD: The effects on inattention,

impulsivity, and hyperactivity: A meta-analysis. *Clinical EEG and Neuroscience, 40*, 180–189.

Aupont, O., Doerfler, L., Connor, D. F., Stille, C., Tisminetzky, M., & McLaughlin, T. J. (2013). A collaborative care model to improve access to pediatric mental health services. *Administration and Policy in Mental Health, 40*(4), 264–273.

Ayers, A. J. (1979). *Sensory integration and the child*. Los Angeles: Western Psychological Services.

Ayers, A. J., & Robbins, J. (2005). *Sensory integration and the child: Understanding hidden sensory challenges*. Los Angeles: Western Psychological Services.

Bagwell, C. L., Molina, B. S. G., Pelham, W. E., & Hoza, B. (2001). Attention-deficit hyperactivity disorder and problems in peer relations: Predictions from childhood to adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(11), 1285–1292.

Baker, A. M. (1980). The efficacy of the Feingold K-P diet: A review of pertinent empirical investigations. *Behavior Disorders, 6*, 32–35.

Bangs, M. E., Jin, L., Zhang, S., Desaiah, D., Allen, A. J., Read, H. A., et al. (2008a). Hepatic events associated with atomoxetine treatment for attention-deficit hyperactivity disorder. *Drug Safety, 31*(4), 345–354.

Bangs, M. E., Tauscher-Wisniewski, S., Polzer, J., Zhang, S., Acharya, N., Desaiah, D., et al. (2008b). Meta-analysis of suicide-related behavior events in patients treated with atomoxetine. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(2), 209.

Barkley, R. A. (2005). *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed.). New York: Guilford Press.

Barkley, R. A. (2012). Effectiveness of restricted elimination diets for management of ADHD: Concerns about the 2011 INCA study. *ADHD Report, 29*(5), 1–5, 11–12.

Barkley, R. A., Copeland, A., & Sivage, C. (1980). A self-control classroom for hyperactive children. *Journal of Autism and Developmental Disorders, 10*, 75–89.

Bateman, B., Warner, J. O., Hutchinson, E., Dean, T., Rowlandson, P., Gant, C., et al. (2004). The effects of a double blind, placebo controlled, artificial food colourings and benzoate preservative challenge on hyperactivity in a general population sample of preschool children. *Archives of Disease in Childhood, 89*(6), 506–511.

Beck, M. H., Cataldo, M., Slifer, K. J., Pulbrook, V., & Guhman, J. K. (2005). Teaching children with attention deficit hyperactivity disorder (ADHD) and autistic disorder (AD) how to swallow pills. *Clinical Pediatrics, 44*(6), 515–526.

Best, J. R. (2012). Exergaming immediately enhances children's executive function. *Developmental Psychology, 48*(5), 1501.

Biederman, J., Baldessarini, R. J., Wright, V., Knee, D., & Harmatz, J. S. (1989a). A double-blind placebo controlled study of desipramine in the treatment of ADD: II. Serum drug levels and cardiovascular findings. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*(6), 903–911.

Biederman, J., Baldesserani, R. J., Wright, V., Knee, D., & Harmatz, J. S. (1989b). A double-blind placebo controlled study of desipramine in the treatment of

ADHD: I. Efficacy. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*(5), 777–784.

Biederman, J., Gastfriend, D. R., & Jellinek, M. S. (1986). Desipramine in the treatment of children with attention deficit disorder. *Journal of Clinical Psychopharmacology, 6*(6), 359–363.

Biederman, J., Melmed, R. D., Patel, A., McBurnett, K., Konow, J., Lyne, A., et al. (2008). A randomized, double-blind, placebo-controlled study of guanfacine extended release in children and adolescents with attention-deficit/hyperactivity disorder. *Pediatrics, 121*(1), e73–e84.

Biederman, J., Spencer, T. J., Newcorn, J. H., Gao, H., Milton, D. R., Feldman, P. D., et al. (2007). Effect of comorbid symptoms of oppositional defiant disorder on responses to atomoxetine in children with ADHD: A meta-analysis of controlled clinical data. *Psychopharmacology, 190*(1), 31–41.

Bird, H. R., Shrout, P. E., Duarte, C. S., Shen, S., Bauermeister, J. J., & Canino, G. (2008). Longitudinal mental health service and medication use for ADHD among Puerto Rican youth in two contexts. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(8), 879–889.

Birmaher, B., & Brent, D. A. (2007). Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(11), 1503–1526.

Bloch, M. H., & Qawasmi, A. (2011). Omega-3 fatty acid supplementation for the treatment of children with attention-deficit/hyperactivity disorder symptomatology: Systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*(10), 991–1000.

Bloomquist, M. L., August, G. J., Cohen, C., Doyle, A., & Everhart, K. (1997). Social problem solving in hyperactive-aggressive children: How and what they think in conditions of automatic and controlled processing. *Journal of Clinical Child Psychology, 26*(2), 172–180.

Boyd, W. D., & Campbell, S. E. (1998). EEG biofeedback in the schools: The use of EEG biofeedback to treat ADHD in a school setting. *Journal of Neurotherapy, 2*, 65–70.

Braswell, L., & Bloomquist, M. L. (1991). *Cognitive-behavioral therapy with ADHD children: Child, family, and school interventions.* New York: Guilford Press.

Bratton, S. C., & Ray, D. (2000). What the research shows about play therapy. *International Journal of Play Therapy, 9*(1), 47.

Bratton, S. C., Ray, D., Rhine, T., & Jones, L. (2005). The efficacy of play therapy with children: A meta-analytic review of treatment outcomes. *Professional Psychology: Research and Practice, 36*(4), 376.

Brown, R. T., Wynne, M. E., & Medenis, R. (1985). Methylphenidate and cognitive therapy: A comparison of treatment approaches with hyperactive boys. *Journal of Abnormal Child Psychology, 13*(1), 69–88.

Buitelaar, J. K., Michelson, D., Danckaerts, M., Gillberg, C., Spencer, T. J., Zuddas, A., et al. (2007). A randomized, double-blind study of continuation treatment for attention-deficit/hyperactivity disorder after 1 year. *Biological Psychiatry, 61*(5), 694–699.

Bussing, R., Zima, B. T., Gary, F. A., & Garvan, C. W. (2002). Use of complementary and alternative medicine for symptoms of attention-deficit hyperactivity disorder. *Psychiatric Services, 53*(9), 1096–1102.

Carlson, G. A., Dunn, D., Kelsey, D., Ruff, D., Ball, S., Ahrbecker, L., et al. (2007). A pilot study for augmenting atomoxetine with methylphenidate: Safety of concomitant therapy in children with attention-deficit/hyperactivity disorder. *Child and Adolescent Psychiatry and Mental Health, 1*(10), 10.

Carper, J. (1998). Miracle cures: Dramatic new scientific discoveries revealing the healing powers of herbs, vitamins, and other natural remedies. Available at *http://lifeplussupplements.securenow.com/link_attention_support.html*.

Chacko, A., Feirsen, N., Bedard, A. C., Marks, D., Uderman, J. Z., & Chimiklis, A. (2013). CogMed working memory training for youth with ADHD: A closer examination of efficacy utilizing evidence-based criteria. *Journal of Clinical Child and Adolescent Psychology, 42*(6), 769–783.

Chai, G., Governale, L., McMahon, A. W., Trinidad, J. P., Staffa, J., & Murphy, D. (2012). Trends of outpatient prescription drug utilization in U.S. children, 2002–2010. *Pediatrics, 130*(1), 23–31.

Childress, A. C. (2012). Guanfacine extended release as adjunctive therapy to psychostimulants in children and adolescents with attention-deficit/hyperactivity disorder. *Advances in Therapy, 29*(5), 385–400.

Childress, A. C., & Sallee, F. R. (2012). The use of lisdexamfetamine dimesylate for the treatment of ADHD. *Expert Review of Neurotherapeutics,12*(1), 13–26.

Conners, C. K. (1980). *Food additives and hyperactive children*. London: Plenum Press.

Conners, C. K., Goyette, C. H., Southwick, D. A., Lees, J. M., & Andrulonis, P. A. (1976). Food additives and hyperkinesis: A controlled double-blind experiment. *Pediatrics, 58*, 154–166.

Connor, D. F., Fletcher, K. E., & Swanson, J. M. (1999). A meta-analysis of clonidine for symptoms of attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*(12), 1551–1559.

Cook, P. S., & Woodhill, J. M. (1976). The Feingold dietary treatment of the hyperkinetic syndrome. *Medical Journal of Australia, 2*, 85–90.

Curtis, L. T., & Patel, K. (2008). *Journal of Alternative and Complementary Medicine, 14*(1), 79–85.

Devlin, S., Healy, O., Leader, G., & Hughes, B. M. (2011). Comparison of behavioral intervention and sensory-integration therapy in the treatment of challenging behavior. *Journal of Autism and Developmental Disorders, 41*(10), 1303–1320.

Diamond, A. (2012). Activities and programs that improve children's executive functioning. *Current Directions in Psychological Science, 21*(5), 335–341.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.

Dujovne, V. F., Barnard, M. U., & Rapoff, M. A. (1995). Pharmacological and cognitive-behavioral approaches in the treatment of childhood depression: A review and critique. *Clinical Psychology Review, 15*(7), 589–611.

DuPaul, G. J., Ervin, R. A., Hook, C. L., & McGoey, K. E. (1998). Peer tutoring for children with attention deficit hyperactivity disorder: Effects on classroom behavior and academic performance. *Journal of Applied Behavior Analysis, 31*(4), 579.

DuPaul, G. J., & Henningson, P. N. (1993). Peer tutoring effects on the classroom performance of children with attention deficit hyperactivity disorder. *School Psychology Review, 22*(1), 134–143.

DuPaul, G. J., & Stoner, G. (1994). *ADHD in the schools: Assessment and intervention strategies*. New York: Guilford Press.

Edwards, L., Salant, V., Howard, V. F., Brougher, J., & McLaughlin, T. F. (1995). Effectiveness of self-management on attentional behavior and reading comprehension for children with attention deficit disorder. *Child and Family Behavior Therapy, 17*(2), 1–17.

Ermer, J., Homolka, R., Martin, P., Buckwalter, M., Purkayastha, J., & Roesch, B. (2010). Lisdexamfetamine dimesylate linear dose-proportionality, low intersubject and intrasubject variability, and safety in an open-label single-dose pharmacokinetic study in healthy adult volunteers. *Journal of Clinical Pharmacology, 50*(9), 1001–1010.

Fabiano, G. A., Pelham, W. E. J., Coles, E. K., Gnagy, E. M., Chronis-Tuscano, A., & O'Connor, B. C. (2009). A meta-analysis of behavioral treatments for attention-deficit/hyperactivity disorder. *Clinical Psychology Review, 29*(2), 129–140.

Faraone, S. V., Spencer, T. J., Kollins, S. H., & Glatt, S. J. (2010). Effects of lisdexamfetamine dimesylate treatment for ADHD on growth. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(1), 24–32.

Fedewa, A. L., & Erwin, H. E. (2011). Stability balls and students with attention and hyperactivity concerns: Implications for on-task and in-seat behavior. *American Journal of Occupational Therapy, 65*, 393–399.

Feingold, B. F. (1973). *Introduction to clinical allergy*. Springfield, IL: Charles C Thomas.

Feingold, B. F. (1975a). Hyperkinesis and learning disabilities linked to artificial food flavors and colors. *American Journal of Nursing, 75*, 797–803.

Feingold, B. F. (1975b). *Why your child is hyperactive*. New York: Random House.

Feingold, B. F. (1976). Hyperkinesis and learning disabilities linked to the ingestion of artificial food colors and flavors. *Journal of Learning Disabilities*, *9*, 19–27.

Findling, R. L., Bukstein, O. G., Melmed, R. D., Lopez, F. A., Sallee, F. R., Arnold, L. E., et al. (2008). A randomized, double-blind, placebo-controlled, parallel-group study of methylphenidate transdermal system in pediatric patients with attention-deficit/hyperactivity disorder. *Journal of Clinical Psychiatry, 69*(1), 149.

Garnock-Jones, K. P., & Keating, G. M. (2009). Atomoxetine: A review of its use in attention-deficit hyperactivity disorder in children and adolescents. *Pediatric Drugs, 11*(3), 203–226.

Geller, D., Donnelly, C., Lopez, F., Rubin, R., Newcorn, J. H., Sutton, V., et al. (2007). Atomoxetine treatment for pediatric patients with attention-deficit/hyperactivity disorder with comorbid anxiety disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(9), 1119–1127.

Gevensleben, H., Holl, B., Albrecht, B., Vogel, C., Schlamp, D., Kratz, O., et al. (2009). Is neurofeedback an efficacious treatment for ADHD?: A randomised controlled clinical trial. *Journal of Child Psychology and Psychiatry, 50*(7), 780–789.

Gilmore, L. (2010). Community knowledge and beliefs about ADHD. *Australian Educational and Developmental Psychologist, 27*(1), 20–30.

Gray, S. A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., et al. (2012). Effects of a computerized working memory training program on

working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 53*(12), 1277–1284.

Green, C. T., Long, D. L., Green, D., Iosif, A. M., Dixon, J. F., Miller, M. R., et al. (2012). Will working memory training generalize to improve off-task behavior in children with attention-deficit/hyperactivity disorder? *Neurotherapeutics, 9*(3), 639–648.

Greenhill, L. L., & Ford, R. E. (2002). Childhood attention deficit hyperactivity disorder: Pharmacological treatments. In P. E. Nathan & J. M. Gorman (Eds.), *A guide to treatments that work* (2nd ed., pp. 25–55). New York: Oxford University Press.

Greenhill, L. L., Pliszka, S., Dulcan, M. K., Bernet, W., Arnold, V., Beitchman, J., et al. (2002). Practice parameters for the use of stimulant medications in the treatment of children, adolescents and adults. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(Suppl. 2), 26S–49S.

Hammerness, P., Georgiopoulos, A., Doyle, R. L., Utzinger, L., Schillinger, M., Martelon, M. K., et al. (2009). An open study of adjunct OROS-methylphenidate in children who are atomoxetine partial responders: II. Tolerability and pharmacokinetics. *Journal of Child and Adolescent Psychopharmacology, 19*(5), 493–499.

Harley, J. P., & Matthews, C. G. (1978). The Feingold hypothesis: Current studies. *Contemporary Nutrition, 3*, 171–173.

Hechtman, L. (2006). Long-term treatment of children and adolescents with attention-deficit/hyperactivity disorder (ADHD). *Current Psychiatry Reports, 8*, 398–408.

Heil, S. H., Holmes, H. W., Bickel, W. K., Higgins, S. T., Badger, G. J., Laws, H. F., et al. (2002). Comparison of the subjective, physiological, and psychomotor effects of atomoxetine and methylphenidate in light drug users. *Drug and Alcohol Dependence, 67*(2), 149–156.

Hinshaw, S. P., Buhrmester, D., & Heller, T. (1989). Anger control in response to verbal provocation: Effects of stimulant medication for boys with ADHD. *Journal of Abnormal Child Psychology, 17*(4), 393–407.

Hinshaw, S. P., Heller, T., & McHale, J. P. (1992). Covert antisocial behavior in boys with attention deficit hyperactivity disorder: External validation and effects of methylphenidate. *Journal of Consulting and Clinical Psychology, 60*, 274–281.

Hinshaw, S. P., Henker, B., & Whalen, C. K. (1984). Self-control in hyperactive boys in anger-inducing situations: Effects of cognitive-behavioral training and of methylphenidate. *Journal of Abnormal Child Psychology, 12*(1), 55–77.

Hinshaw, S. P., Henker, B., Whalen, C. K., Erhardt, D., & Dunnington, R. E. (1989). Aggressive, prosocial, and nonsocial behavior in hyperactive boys: Dose effects of methylphenidate in naturalistic settings. *Journal of Consulting and Clinical Psychology, 57*(5), 636–643.

Hinshaw, S. P., March, J. S., Abikoff, H., Arnold, L. W., Cantwell, D. P., Conners, G. K., et al. (1997). Comprehensive assessment of childhood attention-deficit hyperactivity disorder in the context of a multisite, multimodal clinical trial. *Journal of Attention Disorders, 1*(4), 217–234.

Holborow, P., Elkins, J., & Berry, P. (1981). The effect of the Feingold diet on "normal" school children. *Journal of Learning Disabilities, 14*, 143–147.

Hoover, D. W., & Milich, R. (1994). Effects of sugar ingestion expectancies on mother–child interactions. *Journal of Abnormal Child Psychology, 22*(4), 501–515.

Jain, R., Kollins, S. H., Baily, C., Holloway, W., Horacek, J., & Khayrallah, M. (2008). *Developing a sustained release formulation of clonidine for the treatment of children and adolescents with attention-deficit hyperactivity disorder (ADHD).* Annual meeting of the American College of Neuropsychopharmacology, Scottsdale, AZ.

Jain, R., Segal, S., Kollins, S. H., & Khayrallah, M. (2011). Clonidine extended-release tablets for pediatric patients with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*(2), 171–179.

Jasinski, D. R., & Krishnan, S. (2009). Abuse liability and safety of oral lisdexamfetamine dimesylate in individuals with a history of stimulant abuse. *Journal of Psychopharmacology, 23*(4), 419–427.

Kaduson, H. G. (1997). Play therapy for children with attention-deficit hyperactivity disorder. In H. G. Kaduson, D. M. Congelosi, & C. E. Schaefer (Eds.), *The playing cure: Individualized play therapy for specific childhood disorders* (pp. 197–227). Nothevale, NJ: Aronson.

Kavale, K. A., & Forness, S. R. (1983). Hyperactivity and diet treatment: A meta-analysis of the Feingold hypothesis. *Journal of Learning Disabilities, 16*(6), 324–330.

Kazdin, A. E., Bass, D., Siegal, T. C., & Thomas, C. (1989). Cognitive-behavioral therapy and relationship therapy in the treatment of children referred for antisocial behavior. *Journal of Consulting and Clinical Psychology, 57*(4), 522–535.

Kazdin, A. E., Esveldt-Dawson, K., French, N. H., & Unis, A. S. (1987). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. *Journal of Consulting and Clinical Psychology, 55*(1), 76–85.

Kendall, P. C., & Gosch, E. A. (1994). Cognitive-behavioral interventions. In T. H. Ollendick, N. J. King, & W. Yule (Eds.), *International handbook of phobic and anxiety disorders in children and adolescents* (pp. 415–438). New York: Plenum Press.

Kent, K. M., Pelham, W. E., Molina, B. S. G., Sibley, M. H., Waschbusch, D. A., Yu, J., et al. (2011). The academic experience of male high school students with ADHD. *Journal of Abnormal Child Psychology, 39*(3), 451–462.

Kline, J. P., Brann, C. N., & Loney, B. R. (2002). A cacophony in the brainwaves: A critical appraisal of neurotherapy for attention-deficit disorders. *The Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work, 1*(1), 44–54.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., et al. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(2), 177–186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24*(6), 781–791.

Koenig, K. P., & Rudney, S. G. (2010). Performance challenges for children and

adolescents with difficulty processing and integrating sensory information: A systematic review. *American Journal of Occupational Therapy, 64*(3), 430–442.

Kollins, S. H., Jain, R., Brams, M., Segal, S., Findling, R. L., Wigal, S. B., et al. (2011). Clonidine extended-release tablets as add-on therapy to psychostimulants in children and adolescents with ADHD. *Pediatrics, 127*(6), e1406–e1413.

Krisanaprakornkit, T., Ngamjarus, C., Witoonchart, C., & Piyavhatkul, N. (2010). Meditation therapies for attention-deficit/hyperactivity disorder (ADHD). *Cochrane Database of Systematic Reviews*, *6*.

Kuriyan, A. B., Pelham, W. E., Molina, B. S. G., Waschbusch, D. A., Gnagy, E. M., Sibley, M. H., et al. (2013). Young adult educational and vocational outcomes of children diagnosed with ADHD *Journal of Abnormal Child Psychology, 41*(1), *27*–41.

Lane, S. J., & Schaaf, R. C. (2010). Examining the neuroscience evidence for sensory-driven neuroplasticity: Implications for sensory-based occupational therapy for children and adolescents. *American Journal of Occupational Therapy, 64*(3), 375–390.

Lochman, J. E. (1992). Cognitive-behavioral intervention with aggressive boys: Three-year follow up and preventive effects. *Journal of Consulting and Clinical Psychology, 60*(3), 426–432.

Lochman, J. E., Boxmeyer, C. L., Powell, N. P., Barry, T. D., & Pardini, D. A. (2010). Anger control training for aggressive youth. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 227–242). New York: Guilford Press.

Lofthouse, N., Arnold, L. E., Hersch, S., Hurt, E., & DeBeus, R. (2012). A review of neurofeedback treatment for pediatric ADHD. *Journal of Attention Disorders, 16*(5), 351–372.

Lohr, J. M., Meunier, S. A., Parker, L. M., & Kline, J. R. (2001). Neurotherapy does not qualify as an empirically supported behavioral treatment for psychological disorders. *Behavior Therapist, 24*, 97–104.

Lonigan, C., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology, 27*(2), 138–145.

Lopez, F. A., Childress, A., Adeyi, B., Dirks, B., Babcock, T., Scheckner, B., et al. (2013). ADHD symptom rebound and emotional lability with Lisdexam fetamine Dimensylate in children aged 6 to 12 years. *Journal of Attention Disorders*.

Lubar, J. F., & Shouse, M. N. (1977). Use of biofeedback in the treatment of seizure disorders and hyperactivity. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 1, pp. 203–265). New York: Plenum Press.

Lustig, R. (2012). *Fat chance: The bitter truth about sugar*. London: Fourth Estate.

Marino, L., & Lilienfeld, S. O. (2007). Dolphin-assisted therapy: More flawed data and more flawed conclusions. *Anthrozoos: A Multidisciplinary Journal of the Interactions of People and Animals, 20*(3), 239–249.

Mattes, J. A. (1983). The Feingold diet: A current reappraisal. *Journal of Learning Disabilities, 16*, 319–323.

Mattes, J. A., & Gittleman-Klein, R. (1978). A crossover study of artificial food

colorings in a hyperkinetic child. *American Journal of Psychiatry, 135*, 987–988.

May-Bensen, T. A., & Koomar, J. A. (2010). Systematic review of the research evidence examining the effectiveness of interventions using a sensory integrative approach for children. *American Journal of Occupational Therapy, 64*(3), 403–414.

McCann, D., Barrett, A., Cooper, A., Crumpler, D., Dalen, L., Grimshaw, K., et al. (2007). Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: A randomised, double-blinded, placebo-controlled trial. *Lancet, 370*(9598), 1560–1567.

Meichenbaum, D., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. *Journal of Abnormal Psychology, 77*, 115–126.

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective?: A meta-analytic review *Developmental Psychology, 49*(2), 270–291.

Milich, R., Wolraich, M., & Lindgren, S. (1986). Sugar and hyperactivity: A critical review of empirical findings. *Clinical Psychology Review, 6*, 493–513.

Moen, M. D., & Keam, S. J. (2009). Dexmethylphenidate extended release: A review of its use in the treatment of attention-deficit hyperactivity disorder. *CNS Drugs, 23*(12), 1057–1083.

Molina, B. S. G., Hinshaw, S. P., Swanson, J. M., Arnold, L. E., Vitiello, B., Jensen, P. S., et al. (2009). MTA at 8 years: Prospective follow-up of children treated for combined-type ADHD in a multisite study. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(5), 484–500.

Molina, B. S. G., Pelham, W. E., Gnagy, E. M., Thompson, A. L., & Marshal, M. P. (2007). Attention-deficit/hyperactivity disorder risk for heavy drinking and alcohol use disorder is age specific. *Alcoholism: Clinical and Experimental Research, 31*(4), 643–654.

MTA Cooperative Group. (1999a). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, *56*(12), 1073–1086.

MTA Cooperative Group. (1999b). Moderators and mediators of treatment response for children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, *56*(12), 1088–1096.

Muir, V. J., & Perry, C. M. (2010). Guanfacine extended-release in attention deficit hyperactivity disorder. *Drugs, 70*(13), 1693–1702.

Newcorn, J. H., Kratochvil, C., Allen, A., Casat, C., Ruff, D., Moore, R., et al. (2008). Atomoxetine and osmotically released methylphenidate for the treatment of attention deficit hyperactivity disorder: Acute comparison and differential response. *American Journal of Psychiatry, 165*(6), 721–730.

Newcorn, J. H., Sutton, V. K., Weiss, M. D., & Sumner, C. R. (2009). Clinical responses to atomoxetine in attention-deficit/hyperactivity disorder: The Integrated Data Exploratory Analysis (IDEA) study. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(5), 511–518.

Nigg, J. T., Lewis, K., Edinger, T., & Falk, M. (2012). Meta-analysis of attention-deficit/hyperactivity disorder or attention-deficit/hyperactivity disorder symptoms, restriction diet, and synthetic food color additives. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(1), 86–97.

O'Connor, B., Fabiano, G. A., Waschbusch, D. A., Belin, P. J., Gnagy, E. M.,

Pelham, W. E., et al. (2014). Effects of a summer treatment program on functional sports outcomes in young children with ADHD. *Journal of Abnormal Child Psychology, 42*(6), 1005–1017.

Olson, L. J., & Moulton, H. J. (2004). Use of weighted vests in pediatric occupational therapy practice. *Physical and Occupational Therapy in Pediatrics, 24*(3), 45–60.

O'Neill, S., Rajendran, K., & Halperin, J. M. (2012). More than child's play: The potential benefit of play-based interventions for young children with ADHD. *Expert Review of Neurotherapeutics, 12*(10), 1165–1167.

Panksepp, J. (1998). Attention deficit hyperactivity disorder, psychostimulants, and intolerance of childhood playfulness: A tragedy in the making? *Psychological Science, 7*(3), 91–98.

Panksepp, J., Burgdorf, J., Turner, C., & Gordon, N. (2003). Modeling ADHD-type arousal with unilateral frontal cortex damage in rats and beneficial effects of play therapy. *Brain and Cognition, 52*, 97–105.

Patrick, K. S., Straughn, A. B., Perkins, J. S., & González, M. A. (2009). Evolution of stimulants to treat ADHD: Transdermal methylphenidate. *Human Psychopharmacology: Clinical and Experimental, 24*(1), 1–17.

PBS Frontline (2001, April). *Medicating kids.* Available at *www.pbs.org/wgbh/pages/frontline/shows/medicating.*

Pelham, W. E. (1999a). The NIMH multimodal treatment study for ADHD: Just say yes to drugs? *Clinical Child Psychology Newsletter, 14,* 1–6.

Pelham, W. E. (1999b). The NIMH Multimodal Treatment Study for Attention-Deficit Hyperactivity Disorder: Just say yes to drugs? *Canadian Journal of Psychiatry, 44*(12), 981–990.

Pelham, W. E. (2000). Implications of the MTA study for behavioral and combined treatments. *ADHD Report,* 9–13, 16.

Pelham, W. E., Burrows-MacLean, L., Gnagy, E. M., Fabiano, G. A., Coles, E. K., Wymbs, B. T., et al. (2014). A dose-ranging study of behavioral and pharmacological treatment for children with ADHD. *Journal of Abnormal Child Psychology, 42*(6), 1019–1031.

Pelham, W. E., Carlson, C., Sams, S. E., Vallano, G., Dixon, M. J., & Hoza, B. (1993). Separate and combined effects of methylphenidate and behavior modification on boys with attention deficit-hyperactivity disorder in the classroom. *Journal of Consulting and Clinical Psychology, 61*(3), 506–515.

Pelham, W. E., & Fabiano, G. A. (2008). Evidence-based psychosocial treatment for attention-deficit/hyperactivity disorder: An update. *Journal of Clinical Child and Adolescent Psychology, 37*(1), 184–214.

Pelham, W. E., Foster, E. M., & Robb, J. A. (2007). The economic impact of attention-deficit/hyperactivity disorder in children and adolescents. *Ambulatory Pediatrics, 7*(1S), 121–131.

Pelham, W. E., Gnagy, E. M., Greiner, A. R., & MTA Cooperative Group. (2000). *Parent and teacher satisfaction with treatment and evaluation of effectiveness.* Association for the Advancement of Behavior Therapy, New Orleans, LA.

Pelham, W. E., Hoza, B., Pillow, D. R., Gnagy, E. M., Kipp, H. L., Greiner, A. R., et al. (2002). Effects of methylphenidate and expectancy on children with ADHD: Behavior, academic performance, and attributions in a summer treatment program and regular classroom setting. *Journal of Consulting and Clinical Psychology, 70*(2), 320–335.

Pelham, W. E., Schnedler, R. W., Bender, M. E., Nilsson, D. E., Miller, J., Budrow, M. S., et al. (1988). The combination of behavior therapy and methylphenidate in the treatment of attention deficit disorders: A therpy outcome study. In L. Bloomingdale (Ed.), *Attention deficit disorders: III. New research in attention, treatment, and psychopharmacology* (pp. 29–48). London: Pergamon.

Pelham, W. E., Waschbusch, D. A., Hoza, B., Pillow, D. R., & Gnagy, E. M. (2001). Effects of methylphenidate and expectancy on ADHD boys' performance, self-evaluations, persistence, and attributions on a social task. *Experimental and Clinical Psychopharmacology, 9*(4), 425–437.

Pelham, W. E., Wheeler, T., & Chronis, A. M. (1998). Empirically supported psychosocial treatment for attention deficit hyperactivity disorder. *Journal of Clinical Child Psychology, 27*(2), 190–205.

Pham, T. (2011). *The Effects of Gum Chewing on Classroom Performance in Children with ADHD: A pilot study*. University of Puget Sound.

Plumer, P. J., & Stoner, G. (2005). The relative effects of classwide peer tutoring and peer coaching on the positive social behaviors of children with ADHD. *Journal of Attention Disorders, 9*(1), 290–300.

Polanczyk, G., & Rohde, L. A. (2007). Epidemiology of attention-deficit/hyperactivity disorder across the lifespan. *Current Opinion in Psychiatry, 20*(4), 386–392.

Prince, J. B., Wilens, T. E., Biederman, J., Spencer, T. J., & Wozniak, J. R. (1996). Clonidine for sleep disturbances associated with attention-deficit hyperactivity disorder: A systematic chart review of 62 cases. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*(5), 599–605.

Rabiner, D. L., Murray, D. W., Skinner, A. T., & Malone, P. S. (2010). A randomized trial of two promising computer-based interventions for students with attention difficulties. *Journal of Abnormal Child Psychology, 38*(1), 131–142.

Rapp, D. J. (1978). Does diet affect hyperactivity? *Journal of Learning Disabilities, 11*, 56–62.

Ray, D. C., Armstrong, S. A., Warren, E. S., & Balkin, R. S. (2005). Play therapy practices among elementary school counselors. *Professional School Counseling, 8*(4), 360–365.

Ray, D. C., Schottelkorb, A., & Tsai, M. H. (2007). Play therapy with children exhibiting symptoms of attention deficit hyperactivity disorder. *International Journal of Play Therapy, 16*(2), 95.

Reichenberg-Ullman, J., & Ullman, R. (1996). *Ritalin free kids: Safe and effective homeopathic medicine for ADD and other learning problems*. New York: Crown Publishing Group.

Reid, R., Trout, A. L., & Schartz, M. (2005). Self-regulation interventions for children with attention deficit/hyperactivity disorder. *Exceptional Children, 71*(4), 361–377.

Richters, J. E., Arnold, L. E., Jensen, P. S., Abikoff, H. B., Conners, C. K., Greenhill, L. L., et al. (1995). NIMH collaborative multisite multimodal treatment study of children with ADHD: I. Background and rationale. *Journal of the American Academy of Child and Adolescent Psychiatry, 34*(8), 987–1000.

Riggs, N. R., Greenberg, M. T., Kusché, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. *Prevention Science, 7*(1), 91–102.

Rimland, B. (1983). The Feingold diet: An assessment of the reviews by Mattes, by Kavale and Forness and others. *Journal of Learning Disabilities, 16*, 331–333.

Robb, J. A., Sibley, M. H., Pelham, W. E., Foster, E. M., Molina, B. S. G., Gnagy, E. M., et al. (2011). The estimated annual cost of ADHD to the U.S. education system. *School Mental Health*, *3*(3), 169–177.

Robins, J. (1998, May/June). Wired for miracles? *Psychology Today*, *31*, 41–44.

Rossiter, T. R., & LaVaque, T. J. (1995). A comparison of EEG biofeedback and psychostimulants in treating attention deficit/hyperactivity disorders. *Journal of Neurotherapy, 1*, 48–59.

Rowe, K. S. (1988). Synthetic food colorings and hyperactivity: A double-blind crossover study. *Australian Pediatric Journal, 24*, 143–147.

Rowe, K. S., & Rowe, K. J. (1994). Synthetic food coloring and behavior: A dose-response effect in a double-blind, placebo-controlled, repeated-measures study. *Journal of Pediatrics, 125*, 691–698.

Sallee, F. R., McBurnett, K., & Wigal, T. (2010). *Twenty-four-month effectiveness of guanfacine extended release in children and adolescents aged 6 to 17 years with attention-deficit/hyperactivity disorder.* 65th annual meeting of the Society of Biological Psychiatry, New Orleans, LA.

Scahill, L. (2009). Alpha-2 adrenergic agonists in children with inattention, hyperactivity and impulsiveness. *CNS Drugs, 23*(Suppl. 1), 43–49.

Schwarz, A., & Cohen, S. (2013, March 31). A.D.H.D. seen in 11% of U.S. children as diagnoses rise. *New York Times*.

Seay, B. (1999). *A choreographed campaign of misinformation*. Retrieved August 2, 2013, from *www.skeptictank.org/hs/rit.htm*.

Shaaf, R. C., & Davies, P. L. (2010). Evolution of the sensory integration frame of reference. *American Journal of Occupational Therapy, 64*(3), 363–367.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). CogMed Working Memory Training: Does the evidence support the claims? *Journal of Applied Research in Memory and Cognition, 1*(3), 185–193.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*(4), 628–654.

Sibley, M. H., Pelham, W. E., Molina, B. S. G., Gnagy, E. M., Waschbusch, D. A., Biswas, A., et al. (2011). The delinquency outcomes of boys with ADHD with and without comorbidity. *Journal of Abnormal Child Psychology, 39*(1), 21–32.

Sibley, M. H., Pelham, W. E., Molina, B. S. G., Gnagy, E. M., Waschbusch, D. A., Garefino, A., et al. (2012a). Diagnosing ADHD in adolescence. *Journal of Consulting and Clinical Psychology, 80*(1), 139–150.

Sibley, M. H., Pelham, W. E., Molina, B. S. G., Gnagy, E. M., Waxmonsky, J. G., Waschbusch, D. A., et al. (2012b). When diagnosing ADHD in adults emphasize informant reports, DSM items, and impairment. *Journal of Consulting and Clinical Psychology, 80*(6), 1052–1061.

Silberg, J., Meyer, J., Pickles, A., Simonoff, E., Eaves, L., Hewitt, J., et al. (1996). Heterogeneity among juvenile antisocial behaviours: Findings from the Virginia Twin Study of Adolescent Behavioural Development. In G. R. Bock & J. A. Goode (Eds.), *Genetics of criminal and antisocial behavior* (pp. 76–92). Chicester, UK: Wiley.

Smith, L. (1975). *Your child's behavior chemistry*. New York: Random House.

Spencer, T. J., Newcorn, J. H., Kratochvil, C. J., Ruff, D., Michelson, D., &

Biederman, J. (2005). Effects of atomoxetine on growth after 2-year treatment among pediatric patients with attention-deficit/hyperactivity disorder. *Pediatrics, 116*(1), e74–e80.

Stevens, L. J., Kuczek, T., Burgess, J. R., Hurt, E., & Arnold, L. E. (2011). Dietary sensitivities and ADHD symptoms: Thirty-five years of research. *Clinical Pediatrics, 50*(4), 279–293.

Stine, J. J. (1976). Symptom alleviation in the hyperactive child by dietary modification: A report of two cases. *American Journal of Orthopsychiatry, 46*, 637–647.

Sumner, C. R., Schuh, K. J., Sutton, V. K., Lipetz, R., & Kelsey, D. K. (2006). Placebo-controlled study of the effects of atomoxetine on bladder control in children with nocturnal enuresis. *Journal of Child and Adolescent Psychopharmacology, 16*(6), 699–711.

Swanson, J. M., McBurnett, K., Christian, D. L., & Wigal, T. (1995). Stimulant medications and the treatment of children with ADHD. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical child psychology* (Vol. 17, pp. 265–322). New York: Plenum Press.

Thompson, E., & Ni Bhrolchain, C. (2013). The epidemiology of community paediatric outpatient referrals 2006. *Child: Care, Health and Development, 39*(1), 50–54.

Tinker, R. H., & Wilson, S. A. (1999). *Through the eyes of a child: EMDR with children*. New York: Norton Professional.

Vaughan, B., Fegert, J., & Kratochvil, C. J. (2009). Update on atomoxetine in the treatment of attention-deficit/hyperactivity disorder. *Expert Opinion on Pharmacotherapy, 10*(4), 669–676.

Verbeeck, W., Tuinier, S., & Bekkering, G. E. (2009). Antidepressants in the treatment of adult attention-deficit hyperactivity disorder: A systematic review. *Advances in Therapy, 26*(2), 170–184.

Visser, S. N., Bitsko, R. H., Danielson, M. L., & Perou, R. (2010). Increasing prevalence of parent-reported attention-deficit/hyperactivity disorder among children—United States, 2003 and 2007. *Morbidity and Mortality Weekly Report, 59*(44), 1439–1443.

Wadhwani, S., Radvanski, D. C., & Carmody, D. P. (1998). Neurofeedback training in a case of attention deficit hyperactivity disorder. *Journal of Neurotherapy, 3*, 42–49.

Waschbusch, D. A. (2002). A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. *Psychological Bulletin, 128*(1), 118–150.

Waschbusch, D. A., Cunningham, C. E., Pelham, W. E., Rimas, H. L., Greiner, A. R., Gnagy, E. M., et al. (2011). A discrete choice conjoint experiment to evaluate parent preferences for treatment of young, medication naïve children with ADHD. *Journal of Clinical Child and Adolescent Psychology, 40*(4), 546–561.

Waschbusch, D. A., Fabiano, G. A., & Pelham, W. E. (2012). Evidence-based practice in child and adolescent disorders. In P. Sturmey & M. Hersen (Eds.), *Handbook of evidence-based practice in clinical psychology* (Vol. 1, pp. 27–49). Hoboken, NJ: Wiley.

Waschbusch, D. A., Pelham, W. E., Waxmonsky, J., & Johnston, C. (2009). Are

there placebo effects in the medication treatment of children with ADHD? *Journal of Developmental and Behavioral Pediatrics, 30*(2), 158–168.

Waxmonsky, J. G., Waschbusch, D. A., Akinnusi, O., & Pelham, W. E. (2011). A comparison of atomoxetine administered as once versus twice daily dosing on the school and home functioning of children with attention-deficit/hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology, 21*(1), 21–32.

Waxmonsky, J. G., Waschbusch, D. A., Pelham, W. E., Dragnac-Cordona, L., Rotella, B., & Ryan, L. (2010). Effects of atomoxetine with and without behavior therapy on the school and home functioning of children with ADHD. *Journal of Clinical Psychiatry, 71*(11), 1535–1551.

Wehmeier, P. M., Dittmann, R. W., Schacht, A., Helsberg, K., & Lehmkuhl, G. (2009). Morning and evening behavior in children and adolescents treated with atomoxetine once daily for attention-deficit/hyperactivity disorder (ADHD): Findings from two 24-week, open-label studies. *Child and Adolescent Psychiatry and Mental Health, 3*(1), 5.

Weiss, B. (1982). Food additives and environmental chemicals as sources of childhood behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 21*, 144–152.

Weisz, J. R. (2000, Fall). Lab-clinic differences and what we can do about them: III. National policy matters. *Clinical Child Psychology Newsletter, 15,* 1–4, 6.

Wigal, S. B., Kollins, S. H., Childress, A. C., Squires, L., & 311 Study Group. (2009). A 13-hour laboratory study of lisdexamfetamine dimsylate in school-aged children with attention-deficit/hyperactivity disorder. *Child and Adolescent Psychiatry and Mental Health, 17*(1), 17.

Wigal, T., Brams, M., Gasior, M., Gao, J., Squires, L., & Giblin, J. (2010). Randomized, double-blind, placebo-controlled, crossover study of the efficacy and safety of lisdexamfetamine dimesylate in adults with attention-deficit/hyperactivity disorder: novel findings using a simulated adult workplace environment design. *Behavioral and Brain Functions, 6*(1), 34.

Wilens, T. E., Boellner, S. W., Lopez, F. A., Turnbow, J. M., Wigal, S. B., Childress, A. C., et al. (2008). Varying the wear time of the methylphenidate transdermal system in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(6), 700–708.

Wilens, T. E., Bukstein, O., Brams, M., Cutler, A. J., Childress, A., Rugino, T., et al. (2012). A controlled trial of extended-release guanfacine and psychostimulants for attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(1), 74–85.

Wilens, T. E., Hammerness, P., Utzinger, L., Schillinger, M., Georgiopoulous, A., Doyle, R. L., et al. (2009). An open study of adjunct OROS-methylphenidate in children and adolescents who are atomoxetine partial responders: I. Effectiveness. *Journal of Child and Adolescent Psychopharmacology, 19*(5), 485–492.

Wilens, T. E., Newcorn, J. H., Kratochvil, C. J., Gao, H., Thomason, C. K., Rogers, A. K., et al. (2006). Long-term atomoxetine treatment in adolescents with attention-deficit/hyperactivity disorder. *Journal of Pediatrics, 149*(1), 112–119.

Wolraich, M. L. (1988). Aspartame and behavior in children. In R. J. Wurtman & E. Ritter-Walker (Eds.), *Dietary phenylalanine and brain function* (pp. 201–206). Boston: Birkhauser.

Wolraich, M. L., Lindgren, S. D., Stumbo, P. J., Stegnik, L. D., Applebaum, M. I., & Kiritsy, M. C. (1994). Effects of diets high in sucrose or aspartame on the behavior and cognitive performance of children. *New England Journal of Medicine, 330*, 301–307.

Wolraich, M. L., McKeown, R. E., Visser, S. N., Bard, D., Cuffe, S., Neas, B., et al. (2012). The prevalence of ADHD: Its diagnosis and treatment in four school districts across two states. *Journal of Attention Disorders*.

Wolraich, M. L., Wilson, D. B., & White, J. W. (1995). The effect of sugar on behavior or cognition in children: A meta-analysis. *Journal of the American Medical Association, 274*(20), 1617–1621.

Wymbs, B. T., Molina, B. S. G., Pelham, W. E., Cheong, J., Gnagy, E. M., Belendiuk, K. A., et al. (2011). Risk of intimate partner violence among young adult males with childhood ADHD. *Journal of Attention Disorders, 16*(5), 373–383.

CHAPTER FOURTEEN

The Status of Treatment for Autism Spectrum Disorders

The Weak Relationship of Science to Interventions

Raymond G. Romanczyk, Laura B. Turner, Melina Sevlever, and Jennifer M. Gillis

BACKGROUND AND CONTEXT FOR AUTISM SPECTRUM DISORDERS

Autism is a heterogeneous neurodevelopmental disorder that is currently viewed as part of a spectrum disorder. The diagnostic criteria for what is now termed autism spectrum disorder (ASD) reflect significant changes in DSM-5 (American Psychiatric Association, 2013), as opposed to DSM-IV-TR (American Psychiatric Association, 2000). In DSM-IV-TR the corresponding category was pervasive developmental disorders, and within this category were autistic disorder, Asperger's disorder, pervasive developmental disorder not otherwise specified (PDD-NOS), Rett's disorder, and childhood disintegrative disorder.

Given that DSM-5 adopts the nomenclature of autism spectrum disorders instead of the DSM-IV-TR term "pervasive developmental disorders," for the purposes of this chapter we shall use the term "autism spectrum disorders" to refer to this heterogeneous class of conditions. The following is a brief history and description of ASD with greater detail available elsewhere (e.g., Romanczyk & Callahan, 2012).

History

Leo Kanner (1943) described a syndrome he referred to as "early infantile autism" to emphasize its early onset and observed that the primary characteristic was social "aloneness." He identified a number of specific characteristics, such as:

- Aloofness
- Excellent rote memory
- Echolalia
- Pronoun reversal
- An intense desire to maintain sameness
- Alert expression
- Poor eye contact
- Appearance of being deaf
- No anticipatory reaching out from infancy
- Lack of social and communication initiation
- Normal intelligence (as assessed by the Seguin Form Board)
- Normal motor coordination

Kanner also wrote that he considered the "essential common characteristic:" of individuals with autism to be an "inability to relate themselves in the ordinary way to people and situations from the beginning of life" (p. 41). This emphasis on aloofness, as Kanner referred to it, has remained the primary feature of the disorder across DSM editions. The DSM-5 criteria now emphasize this deficit by highlighting the longstanding difficulties in interpersonal communication and interaction observed across multiple settings, which cannot be explained by more general delays in development.

Diagnostic History Up to DSM-5

The third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III; American Psychiatric Association, 1980) first listed autism as a disorder. It created a new class of disorders—the pervasive developmental disorders—and infantile autism was placed within this group. DSM-III required six diagnostic criteria, including emergence of symptoms prior to two and a half years of age; severe and situationally consistent lack of responsiveness to others; serious disturbances in linguistic development; echolalia, reversal of pronouns, and other speech abnormalities (when speech is present); strange responses to stimuli, such as extreme resistance to even minor changes in the environment or preoccupation with mechanical objects; and the absence of psychotic features typical of schizophrenia (e.g., delusions).

The revision to DSM-III, DSM-III-R (American Psychiatric Association, 1987), introduced the term "autistic disorder." The diagnostic criteria

were modified significantly, yielding three broad categories, namely, serious deficits in back-and-forth interactions with others; serious deficits in communication, both linguistic and nonlinguistic (e.g., eye contact); and extremely restricted range of interests and behaviors (e.g., fascination with train schedules). Within these three overarching categories, 16 specific behaviors were listed. The diagnosis of autistic disorder required at least half of the 16 behaviors to be present, but a distribution requirement was imposed across the three categories: at least two from the reciprocal social interaction category and at least one from each of the other two categories. The criteria were further modified in 1994 in DSM-IV, and in 2000 in DSM-IV-TR, the three categories and specific distribution requirements were retained. Interestingly, adoption of the DSM-IV criteria was associated with an increase in cases of autistic disorder compared with the time period in which DSM-III criteria were used, reflecting the broadening of the criteria in DSM-IV-TR compared with DSM-III.

DSM-5 (American Psychiatric Association, 2013) adopted major changes to the conceptualization and diagnosis of autism. The first major change concerns the organization of pervasive developmental disorders (defined earlier), a class of disorders eliminated from DSM-5. Autism spectrum disorder is now a new diagnosis within the DSM-5 class of neurodevelopmental disorders. This change was inspired, in part, by the problematic reliability of diagnoses among the disorders that comprise pervasive developmental disorders. There is extreme variability in the prevalence of the diagnoses within the pervasive developmental disorders across diagnostic centers, and low reliability among qualified professionals (Lord et al., 2011). Low reliability may stem in part from the reliance on a categorical model, which may lead to discrepancies among clinicians who disagree on only one diagnostic criterion. These findings support Lord and colleagues' (2011) recommendation to modify the current categorical model of related disorders to a dimensional (i.e., spectrum) model with descriptions of core features.

Consistent with current research on autism, the new diagnostic criteria for autism spectrum disorder reduce the old three-category model to a two-category model, reassigning most of the symptoms associated with impairments in verbal and nonverbal communication to either of the following two categories: (1) persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays; and (2) restricted, repetitive patterns of behavior, interests, or activities. Additionally, given that delays in the development of spoken language are characteristic of several childhood disorders and not exclusive to autism, this symptom is no longer a diagnostic criterion. However, several abilities, such as expressive language and intellectual functioning, have now been conceptualized and described on a dimension accompanying an individual's diagnosis of ASD.

The specific distribution requirements have also been modified,

requiring two or more symptoms than did DSM-IV-TR. In DSM-5, individuals must exhibit all three symptoms associated with a persistent deficit in social communication and social interaction and at least two symptoms characteristic of restricted, repetitive patterns of behavior, interests, or activities. In contrast to Lord and colleagues (2011), some experts in the field (e.g., Dawson, 2012; McPartland, Reichow, & Volkmar, 2012) have expressed concerns that the DSM-5 criteria are too strict and that some individuals will not receive a diagnosis and thereby not obtain needed educational or psychological services. However, extant prospective data do not permit conclusive statements about the impact of the DSM-5 modifications on the diagnosis and treatment services for individuals with an ASD.

Heterogeneity and Comorbidity

Heterogeneity comes in many forms, ranging from variability along a dimension of a specific symptom to variability within the makeup of symptom clusters, to variability within the range and severity of the symptoms of a diagnosis. ASD presents a particular challenge because all of these factors contribute to the complexity and diversity of the individuals who receive a diagnosis. More specifically, DSM-IV-TR listed 12 symptoms for autistic disorder but required only six symptoms, distributed among the three categories to obtain a diagnosis. This leaves significant room for heterogeneous symptom clusters among individuals with the disorder. Further, developmental variables, such as language level and intellectual functioning, can affect the topography and severity of autism-specific symptoms, such as social communication deficits and restricted, repetitive behaviors (see Lord & Jones, 2012, for a review).

The presence of comorbid psychiatric conditions adds further complexity. Diagnostic criteria for autism have historically been exclusionary in that a diagnosis of autism precluded the diagnosis of other psychiatric disorders. However, given changes to the diagnostic criteria over the years, individuals with ASD are now commonly diagnosed with a second or third psychiatric disorder. Results of a recent study indicated that 70% of a sample of 112 10- to 14-year-old children and adolescents with an ASD had at least one comorbid psychiatric disorder and 41% had at least two (Simonoff et al., 2008). Common comorbid diagnoses include several anxiety disorders, depression, attention-deficit/hyperactivity disorder (ADHD), sleep disorders, and conduct and oppositional defiant disorders (Bradley, Summers, Wood, & Bryson, 2004; Ghaziuddin, Weidmer-Mikhail, & Ghaziuddin, 1998; Simonoff et al., 2008). Additionally, up to 75% of individuals with an ASD show some level of intellectual disability (Croen, Grether, & Selvin, 2002).

Thus, it is not surprising that the fields of psychology and psychiatry, not to mention other fields and the general public, have a wide range of opinions with respect to assessment and treatment of ASD. The heterogeneity

within and between disorders comprising ASD that are often clustered for descriptive purposes in news media, service delivery, advocacy, research, and public policy creates a fertile context for pseudoscience.

Prevalence

Since the addition of autism to DSM-III in 1980, the prevalence of autism has leapt from 4.5 per 10,000 to 1 in 88 for the target age of 8 years old (Centers for Disease Control and Prevention, 2012) and 1 in 50 for school-age children ages 6–17 (Centers for Disease Control and Prevention, 2013). Several explanations have been offered for this increase (King & Bearman, 2009). The reason that is perhaps receiving the most attention is the change in diagnostic criteria, as the prevalence of autism has markedly increased since DSM-IV and has continued to increase since DSM-IV-TR. It is unclear what, if any, changes in prevalence will result from the changes in DSM-5.

Public awareness of ASD has also increased. Romanczyk and Callahan (2012) presented a telling analysis of a representative popular press outlet, *Time* magazine. In the 1980s, three articles were published on autism; in the 1990s, this number increased to 11 articles and two cover stories; and in the 2000s, this number increased further to 73 articles and six cover stories. Currently, there are similar high levels of emphasis in print and television media, including frequent public service announcements. Along with an increase in public awareness of the disorder, in the last decade early identification of ASDs has been mandated by organizations such as the American Academy of Pediatrics. The organizations suggest that pediatricians screen for autism and monitor social-emotional developmental multiple times during the first 3 years of age (Myers & Johnson, 2007), increasing the likelihood that children on the spectrum will be identified.

Etiological Complexity

The specific causes of autism are not yet known. However, there are multiple hypotheses for its etiology based on research in neurobiology, genetics, and psychology. Given the heterogeneity of autism, there are probably multiple contributors to its development. The lack of specific causes, as well as likely multiple pathways to its genesis, may support radical and at times pseudoscientific explanations.

A prime example is the widespread assumption that the measles, mumps, rubella (MMR) vaccine causes autism. Several issues are intertwined—the MMR vaccine itself, the scheduling/distribution of administration of child vaccinations, and the use of thimerosal (which contains ethylmercury) as a preservative in vaccines. In 1998, British pediatrician Andrew Wakefield and colleagues published a paper in *The Lancet*, which purported to find a causal relation between the MMR vaccine, autism, and gastrointestinal disorders (Wakefield et al., 1998). As a consequence, many parents ceased

vaccinating their children. For years, other researchers were unable to replicate this finding, leading to an investigation, which concluded that Wakefield had falsified the medical histories of the 12 children in his 1998 study. *The Lancet* immediately retracted the publication (The Editors of *The Lancet*, 2010), and Wakefield has since lost his medical license. Despite this retraction, many parents remain cautious about vaccinating their children.

In 1999, out of an abundance of caution, the American Academy of Pediatrics and the U.S. Public Health Service required thimerosal to be removed from childhood vaccines, including the MMR vaccine. To track the potential impact on the rate of ASD, Schechter and Grether (2008) analyzed the prevalence of ASD in California from 1995 to 2007. If thimerosal were a causal factor in the development of autism, the results should have pointed to a decrease in prevalence. Yet, there was no decrease in the prevalence of autism following the removal of thimerosal from vaccines. Moreover, recent epidemiological research suggests no link between vaccines and autism (DeStefano, Price, & Weintraub, 2013).

EFFICACIOUS TREATMENT

The Research Base

The research base concerning ASD is extensive and diverse. For the purpose of this section, we focus on well-conducted and consistently replicated studies to serve as a basis of comparison for popular interventions in the following section.

The seminal study in this vein was the UCLA Young Autism Project (Lovaas, 1987). In this investigation, Lovaas evaluated the educational effects of a 2-year, behaviorally based intervention for young children (i.e., under 46 months of age) diagnosed with autism. Children were assigned to one of two groups, differing only in the intensity of the intervention. The first group received at least 40 hours/week of 1:1 adult-to-child ratio behavioral intervention (i.e., "intensive intervention"; $n = 19$), whereas the second group received no more than 10 hours/week (i.e., "minimal intervention"; $n = 19$). Outcome measures were collected when the children were 6–7 years old and indicated that the group receiving intensive behavioral intervention scored significantly higher in educational placement and intellectual functioning than the group receiving minimal behavioral intervention. More specifically, nine of the children receiving intensive intervention passed first grade (as compared with one child in the minimal intervention group), and the intensive intervention group had a mean IQ score on average 30 standard points higher than the group receiving minimal behavioral intervention.

A few years later, McEachin, Smith, and Lovaas (1993) followed up the children from the Lovaas (1987) study to evaluate whether the intensive intervention group maintained their gains. The authors collected measures of educational placement, intellectual functioning, and adaptive behavior

for children in both the intensive and minimal intervention groups when the children were on average 13 and 10 years old, respectively. Results for the intensive intervention group indicated that, out of the original 19, eight were in typical education classes, and 11 had an IQ of at least 80, suggesting nearly half had obtained "best outcomes." In contrast, results for the minimal intervention group (*n* also of 19) indicated that no child was in typical education classes and only three children received IQ scores of at least 80. Additionally, the group receiving the intensive behavioral intervention achieved significantly higher adaptive behavior scores than the group receiving minimal behavioral intervention.

Using archival data, Smith, Eikeseth, Klevstrand, and Lovaas (1997) conducted one of the first replications of the UCLA Young Autism Project. The authors examined outcomes of a 2-year intensive behavioral intervention in preschoolers with severe intellectual disability and a pervasive developmental disorder (PDD). Similar to the Lovaas (1987) study, children were assigned to either receive intensive (i.e., 30 hours/week; $n = 11$) or minimal (i.e., 10 hours/week; $n = 10$) behavioral intervention. Following the intervention, although no children performed in the average range of intellectual functioning, the mean IQ for children in the intensive group was significantly higher than was the mean IQ for the minimal group. Outcome data also indicated that approximately 90% of the children in the intensive group began using words, phrases, and/or sentences, whereas only 20% of the children in the minimal group were using single words.

Sheinkopf and Siegel (1998) expanded the previous findings by evaluating an intensive, 16-month, home-based intensive behavioral intervention in children diagnosed with autistic disorder or PDD-NOS. The intensive intervention group received home-based behavioral intervention for an average of 27 hours/week. The control group received "services as usual" for 11 hours/week. At follow-up, children receiving the intensive behavioral intervention earned significantly higher IQ scores than children in the control group, with a mean difference of approximately 25 points.

Despite the growing body of outcome data to support the efficacy of intensive behavioral intervention, "eclectic" interventions are widely available to children with an ASD in the public school systems. Accordingly, Howard, Sparkman, Cohen, Green, and Stanislaw (2005) compared the effects of three treatment approaches: (1) intensive behavioral intervention (1:1 adult-to-child ratio, 25–40 hours per week); (2) intensive "eclectic" intervention (a combination of methods, 1:1 or 1:2 ratio, 30 hours per week); and (3) nonintensive public early intervention programs (a combination of methods, small-group instruction, 15 hours per week). At intake, children in the three groups obtained similar scores on measures of intellectual, language and adaptive functioning. At follow-up, the children receiving intensive behavioral intervention displayed superior performance in all skill domains relative to those in the other two groups. Not surprisingly, there were no statistically significant differences between the intensive "eclectic" and nonintensive public early interventions.

In an effort to systematically summarize the findings of the aforementioned studies, along with the findings of several similar outcome studies, Eikeseith (2009) evaluated 25 studies that utilized a comprehensive psychoeducational approach for early intervention. Twenty studies evaluated behavioral intervention, and five studies evaluated well-known broad educational programs. Outcome studies were evaluated for scientific value and effect size of results. The authors concluded that according to the guidelines established by Chambless et al. (1996) and Chambless and Hollon (1998), only applied behavior analysis (ABA) is "well established," which is considered to be the highest level of scientific support. (Note: The terms "intensive behavioral intervention" [IBI] and "applied behavior analysis" are often used interchangeably.)

Furthermore, Eldevik et al. (2009) conducted a meta-analysis of 34 studies of behavioral interventions. Results supported the use of intensive behavioral intervention as the treatment of choice based on superior positive outcomes found reliably across studies. Following up on these results, Eldevik, Hastings, Hughes, Jahr, and Eikeseth (2010) aggregated data from 16 group-design treatment studies. A total of 309 children received intensive behavioral intervention, 39 children received other interventions (i.e., comparison groups), and 105 children participated in a no-treatment control group. For 29.8% of the children receiving intensive behavioral intervention, IQ increased substantially. In contrast, similar gains in IQ were reported for only 2.6% and 8.7% of participants in the comparison and control groups, respectively. A similar pattern of results emerged for changes in adaptive behavior.

In summary, the research literature shows strong support for intensive behavioral intervention in the treatment of children with ASD. This evidence base is compelling in its consistency and scope. The next section provides an overview of practice guidelines that were developed from these and hundreds of related findings.

Practice Guidelines

Practice guidelines are quite different from standard research literature reviews and are intended to provide evidence-based guidance to assist consumers, providers, and regulators in making informed treatment choices for specific disorders or conditions. There are several types of practice guidelines:

- Assessment/intervention guidelines (such as managing otitis media in children).
- Risk assessment/prevention guidelines (such as risks from drug use in pregnancy).
- Administrative guidelines (such as insurance preapproval for surgery).

The Institute of Medicine (IOM) (Field & Lohr, 1992) produced criteria for when clinical practice guidelines are needed:

- The problem is common or expensive.
- Great variation exists in practice patterns.
- There is enough scientific evidence to determine appropriate / optimal practice.

The methodology for creating practice guidelines was initially developed by the Agency for Health Care Policy and Research (AHCPR), established in 1997; the AHCPR is currently known as the Agency for Healthcare Research and Quality (AHRQ) and is part of the United States Public Health Service. It is the federal agency involved with health services research.

As a result of these agencies' efforts and the work of many individuals, there is an accepted methodology for evaluating treatments to produce best practice guidelines (Noyes-Grosser et al., 2005). This methodology has only recently been applied to ASD, with the New York State Department of Health (NYSDOH) Early Intervention Program (EIP) being the first to use this methodology to develop a series of evidence-based clinical practice guidelines.

When examining review projects that use the type of methodology listed earlier, the results are quite consistent: Approaches termed intensive behavioral intervention and applied behavior analysis, within the broad family of behavioral oriented interventions, produce the most significant and consistent results. A sample of such reviews include:

- New York State Department of Health, Early Intervention Program. (1999). *Clinical practice guideline: Guideline technical report. Autism/pervasive developmental disorders, assessment and intervention for young children (ages 0–3 years)*, no. 4217. Albany, NY: New York State Department of Health.
- National Research Council. (2001). *Educating children with autism*. Washington, DC: National Academy Press.
- National Autism Center. (2009). *National standards report: National standards project—Addressing the need for evidence-based practice guidelines for autism spectrum disorders*. Randolph, MA: National Autism Center.
- Missouri Department of Mental Health. (2012). *Autism spectrum disorders: Guide to evidence-based interventions*. Available at *www.autismguidelines.dmh.mo.gov*.

Despite the extensive literature base and available guidelines, there remains general disarray concerning intervention recommendations and preferences in the field of ASD. Clearly, well-conducted research studies,

academic literature reviews, and practice guidelines have been insufficient to promote the widespread acceptance of empirically supported interventions, as hundreds of nonempirically based treatments continue to be widely used (Green et al., 2005). Romanczyk, Gillis, White, and DiGennaro (2008) presented an analysis of over 400 purported treatments culled from a sample of websites and found that less than 1% are supported by outcome research. The next section describes a small sampling of such non–evidence-based interventions.

NONEFFICACIOUS TREATMENTS

Vitamin B_6

Linus Pauling (1968) introduced orthomolecular psychiatry, an approach to treating mental illness using vitamins and minerals. Pauling argued that inadequate concentrations of vitamins and minerals can produce mental illness. His theory has been used to support vitamin treatment for a variety of mental illnesses, including ASD. The use of vitamins to treat many psychological and developmental disorders (e.g., Down syndrome, fragile-X syndrome, and ADHD) is unsupported. Moreover, vitamin therapy has shown no effects for ASD in well-designed studies (Kozlowski, 1992).

Vitamin B_6 is the most studied and controversial vitamin treatment for individuals with ASD. The vitamin is a particularly plausible treatment for ASD, as it is involved in the formation of several neurotransmitters, including serotonin, aminobutyric acid, dopamine, norepinephrine, and epinephrine (Pfeiffer, Norton, Nelson, & Shott, 1995). Evidence that suggested individuals with an ASD have difficulty synthesizing these neurotransmitters legitimized the use of vitamin B_6 as a treatment for autism (Pfeiffer et al., 1995).

According to a recent parent survey, vitamins were the second most commonly used complementary and alternative treatment for children with ASD. Approximately 16–20% of parents report using vitamin B as a treatment for ASD (Christon, Mackintosh, & Myers, 2010; Hall & Riccio, 2012). Moreover, a recent survey of physicians found that 49% of physicians encouraged the use of vitamins for ASD (Golnik & Ireland, 2009). Vitamin B_6 is a collective term for three naturally occurring pyradines: pyridoxine, pyridoxal, and pyridoxamine. In the field of ASD research, vitamin B_6 is referred to as pyridoxine and is typically administered with magnesium. The observation that large doses of vitamin B_6 administered alone can produce undesirable effects (e.g., irritability, hypersensitivity to sound, and enuresis) led to the now standard practice of co-administering magnesium, which was hypothesized to counter these effects (Nye & Brice, 2009).

Bernard Rimland and Gilbert Lelord and colleagues are largely responsible for the early research on the behavioral and physical effects of vitamin

B_6. These researchers typically reported improvements in individuals with autism after vitamin B_6 treatment. However, the majority of studies they conducted possess methodological shortcomings (see Pfeiffer et al., 1995, for a review of these studies), including parent report questionnaires with questionable reliability and validity, infrequent measurement, and questionable clinical significance of findings.

Nye and Brice (2009) conducted an updated literature review and identified 17 published studies examining the effectiveness of B_6–magnesium treatment for autism. Of these studies, only three used adequate control procedures (e.g., random assignment to experimental and control conditions and concealment of assignment from participants; Findling et al., 1997; Kuriyama et al., 2002; Tolbert, Haigler, Waits, & Dennis, 1993). Only two studies reported using double-blind procedures (Findling et al., 1997, and Kuriyama et al., 2002). We describe studies that incorporated control procedures below.

Tolbert et al. (1993) and Findling et al. (1997) reported no observed improvements after B_6–magnesium treatment. However, these studies contained methodological flaws that complicate their interpretation. Tolbert and colleagues (1993) conducted an open trial and administered lower doses of B_6–magnesium combinations than typically used. The authors found no difference in behavioral ratings using unstandardized measures. Findling and colleagues (1997) conducted a double-blind, placebo-controlled trial but did not use standardized measures. They found no difference between experimental and control groups, but did find evidence for a placebo effect in vitamin treatments. The largest change on the Childhood Psychiatric Rating Scale (CPRS) occurred in the first 2 weeks, when all subjects were placed on placebo pills.

Kuriyama and colleagues (2002) conducted a double-blind randomized control trial of pyridoxine. Four participants with PDDs were placed on a 4-week trial, and four participants matched in IQ and age were placed on a placebo. The authors reported a significant increase in verbal IQ (VIQ) for the pyridoxine group (11 points); however, this increase was only 5.2 points greater than increases in VIQ found for the control group. Additionally, at baseline the mean VIQ scores for the groups differed by only 3 points. Independent analyses conducted by Nye and Brice (2009) of Kuriyama et al.'s results revealed no significant differences in scores across groups. Notably, a difference of 5 points on the intelligence test used by the authors (Wechsler Intelligence Scale for Children–III) is within the margin of error. Moreover, the authors found no difference in performance IQ (PIQ) scores across groups.

The only large (n = 30) double-blind, placebo-controlled published clinical trial of vitamin B treatment found the vitamin to be ineffective for the behavioral symptoms of ASD. Notably, because the authors examined the effects of vitamin B_{12}, the results may not be generalizable to vitamin B_6. Bertogilo, James, Deprety, Brule, and Hendren (2011) assessed

behavioral changes using standardized, valid, and reliable assessment measures (e.g., the Child Behavior Checklist, the Childhood Autism Rating Scale, the Peabody Picture Vocabulary Test, 3rd ed.). The authors identified no significant differences between active and placebo groups across their behavioral measures. Bertogilo et al. (2011) reported that a subset of nine children appeared to respond positively to the treatment, as indicated by clinically significant increases on a clinician measure of improvement and at least two behavioral measures. Accordingly, the authors tentatively suggested that vitamin B_{12} may yield positive outcomes for a subset of the ASD population, despite the finding that mean outcomes across the treatment and placebo groups did not differ.

The widespread use of vitamin therapy among individuals with ASD is concerning given the lack of support for this treatment and the known side effects of megadoses of many vitamins, including B_6. Megadoses of vitamins C and B_3 have been shown to have direct toxicity, and side effects have been reported for vitamin B_6. The negative side effects of vitamin B_6 megadoses alone include possible physical dependence and withdrawal symptoms (e.g., seizures) when B_6 use is terminated.

Floortime (DIR)

The developmental, individual-difference, relationship-based model (DIR) is described as a "functional developmental approach" that "examines how children integrate their capacities (motor, cognitive, language, spatial, sensory) to carry out emotionally meaningful goals" (Greenspan & Wieder, 1999, p. 148). DIR is used to treat children with a variety of developmental problems, including children with ASD. The approach, sometimes referred to as "Floortime," remains popular with families of children with ASD through publications, seminars, and Internet sites.

In the book *The Child with Special Needs,* Greenspan and Wieder (1998) outlined their conceptual model of the causes, assessment, and treatment of children with ASD. The model is based on Greenspan and Wieder's (1999) affect–diathesis hypothesis, which proposes that children with symptoms of autism may have a "unique biologically based processing deficit involving the connection of affect, motor planning, and sequencing capacities, and symbol formation" (p. 150). According to the authors, the ability to connect affect to motor function and symbolic representation is a critical skill that develops in the second year of life.

The affect–diathesis hypothesis assumes that affect is central to social and cognitive development. Thus, the child's emotional development is considered a core developmental process that is the basis for all skills. For example, a child may learn cause and effect through noninteractional (e.g., accidentally dropping a spoon and hearing a "clang") or emotional (e.g., smiling at mom and having mom return the smile) means. Greenspan and Wieder (1998) hypothesized that a child who learns cause and effect in an

emotional context will develop a stronger foundation in cause-and-effect learning than other children. Thus, Greenspan's theory suggests that "the core psychological deficit in autism" is "an inability to connect emotions and intent to motor planning and sequencing and to emerging symbols" (Greenspan & Shanker, 2004, p. 299).

The complex DIR model presumes that children with ASD possess underlying sensory processing and/or motor planning difficulties that hinder cognitive and affective development. The model also emphasizes the fit of a family's interactional pattern with the child's specific deficits. For example, if a father has a loud expressive style with a child who displays auditory hypersensitivity, the child is likely to have more trouble overcoming his or her sensory difficulties. According to the model, the discrepancy in fit would lead the child to retreat further into a world of isolation to escape the overstimulation of the family environment. Thus, the model emphasizes identifying the child's biological weaknesses (e.g., sensory-processing difficulties, motor planning and sequencing problems) and the family's interactional patterns to design an individualized treatment plan (Greenspan & Wieder, 1998, 2000).

The Floortime technique is a method for remediating the child's biological and developmental needs. Floortime is a child-directed, daily playtime described by Wieder and Greenspan (2003) as "the component that is spontaneous and led by the child, where the caregiver follows the child's lead and promotes the continuous flow of interactions utilizing affect cues that entice, challenge, soothe and encourage the child further" (p. 427). The goals of Floortime are to establish and stabilize six fundamental developmental milestones: (1) self-regulation and interest in the world, (2) intimacy, (3) two-way communication, (4) complex communication, (5) emotional ideas, and (6) emotional thinking. According to the authors, attainment of each of these milestones occurs in a stage-like progression. Although children with autism or PDD may possess splinter skills (i.e., specialized domains of cognitive ability) with respect to each of the milestones, mastery of each of the milestones is considered necessary for typical development. The Floortime technique purports to teach parents how to help their child sequentially master the milestones through relationship-building play. Moreover, the technique is believed to help some children become "more trusting," "intimately related to parents," and "joyful" (Greenspan & Wieder, 1998, p. 463).

Although Greenspan and Wieder's (1998) model provides some useful guidance for clinical decision making, including individualized approaches to assessment and treatment, a focus on motivational variables, and expected developmental progression, it lacks a scientific foundation. Many aspects of the theory appear to be based on clinical experience and anecdotal evidence. The assumptions regarding the powerful role of emotion in cognitive development are not supported by developmental, learning, or cognitive research. As acknowledged by Greenspan and Wieder (1998),

"That the emotional lesson comes first and is the basis for the cognitive lesson is opposite to the traditional view of cognition and learning" (p. 123).

Greenspan and Wieder (1997; in Greenspan & Wieder, 1998) conducted a chart review of 200 children with autism or PDD that is often cited as support for the DIR model. Besides lacking experimental control, the study contained many of the classic flaws of such research, including a biased sample, use of unvalidated outcome measures, and use of a single clinician who was not blind to the hypotheses prior to review progress on the charts.

More recently, Wieder and Greenspan (2005) conducted a follow-up examination of a subgroup of 16 children from the original chart review who were described as "good to outstanding" responders to the intervention. The study is replete with methodological problems that limit interpretation of the findings. As in the original study, the follow-up study lacked a control group and failed to control for the effect of maturation. Moreover, the study relied on a biased sample of participants who were originally classified as "responders." Despite the use of a variety of questionnaires and rating scales, the study provided descriptive, rather than objective, data. For example, the study was largely based on interviews of the boys and their families conducted by parents and the first and second author. The authors advanced excessive and unsubstantiated claims regarding increased empathy (suggested in some cases to be higher than in typically developing children) and appropriate theory of mind skills (the method of assessment of these skills was not described). Despite the lack of systematic evidence, Wieder and Greenspan (2005) reported that the adolescents "became warm, related, and sensitive young people who have the foundations for an optimistic future" who "did not evidence the deficits or symptoms of ASD" (p. 59). Of note, no standardized instruments of ASD symptoms (e.g., the CARS, ADOS, ADI-R) were administered. Moreover, although the authors attribute gains to participation in DIR, they conceded that participants were enrolled in an average of eight treatments, including speech and language therapy, occupational therapy, and special education services, rendering multiple treatment interference a plausible rival hypothesis.

Wieder and Greenspan (2003) also published a more descriptive case report of a 3-year-old child. The boy participated in DIR with his father over 4 years. The comprehensive program included "six daily floor time sessions, four semi-structured and sensory–motor activities, intensive speech and occupational therapies, three to five play dates weekly, inclusion in preschool, and various music, gym, drama, and sports activities" (p. 430). The authors relied on anecdotal and subjective clinical opinion to suggest improvements in the child's social engagement and cognitive skills. In addition to the methodological flaws inherent in any single case report (e.g., lack of a control group, maturation effects), the authors neglected to use any sources of objective data (e.g., the use of standardized parent and

clinician rating scales, standardized measurements of IQ, language, and/or motor skills).

Hilton and Seal (2007) administered a brief clinical trial to examine differences in DIR and applied behavior analysis (ABA). The participants were 2-year-old monozygotic twin brothers with severe language delays and diagnoses of autism. The twins were randomly assigned to receive the DIR or ABA intervention. One-hour sessions were conducted twice a week for a total of 14 sessions. Notably, the delivery of ABA services is recommended to be intensive (40 hours per week); thus, the present intervention delivered a much lower proportion of ABA services than is typically advised. The authors assessed progress during the initial and final session using the Communication and Symbolic Behavior Scales (CSBS). At pretest the boys differed widely on the measure (standard score of 7 and 12). Overall, the child who completed the ABA trial improved from a standard score of 7 to 8, whereas the child who completed the DIR child dropped from a score of 12 to 10. The authors noted that gains and losses were evident across the different clusters assessed within the scale (e.g., gestural and vocal communication, social-affective signaling, reciprocity). The child who participated in ABA showed a greater improvement in response to names, following one-step directions, and using signs. The child who completed DIR showed greater improvement in only imitative and spontaneous production of words. The small sample size, the use of a clinician-rated scale (completed by a clinician who was not blind to treatment status), the inability to match participants at pretest, and low dosage of treatment sessions complicate the interpretation of these results.

Only one quasi-experimental design examining the efficacy of DIR for children with ASD has been conducted. Pajareya and Nopmaneejumruslers (2011) conducted a pilot study in Thailand to examine the effect of adding DIR treatment to standard 20–40 hours per week of intensive ABA intervention. Children were randomly assigned to receive either intensive ABA services alone or intensive ABA services and supplemental DIR treatment implemented by trained parents for an average of 15 hours per week. The authors found that participants who received DIR showed significantly greater decreases on the Childhood Autism Rating Scale (CARS) and two measures developed to assess DIR progress (the Functional Emotional Assessment Scale and the Functional Emotional Developmental Questionnaires). However, although the difference in CARS scores across groups was statistically significant, it is not clinically significant (average difference in means = 2.1). Furthermore, the results are compromised by a lack of researcher blindness to treatment condition.

The DIR approach appears attractive to parents because of its focus on the atypical emotional reciprocity characteristic of autism. The model provides hope to parents that "normal" social interactions can be achieved only with the DIR approach because it focuses on emotional development. By accusing behavioral interventions, albeit falsely, of not focusing on

emotional development, advocates of DIR misrepresent the approach as unique in its ability to provide parents with a warm relationship with their child with autism.

Facilitated Communication

Social and communicative impairments comprise the core characteristics of autism. In this context of dual impairment, facilitated communication (FC) quickly became a nationwide phenomenon in the treatment of autism, with features on prominent television programs, such as *ABC Prime Time Live* and *PBS Frontline* (Kaplan, 1992; Palfreman, 1993). The procedure was rapidly adopted by mainstream clinicians and school programs. In response to the tremendous increase of interest in FC, a special institute for FC at Syracuse University was created (the program continues to exist today, although the term "supported typing" has replaced the term "FC"), training seminars were conducted across the country, and extravagant and unwarranted claims of FC's success were reported by the media. Although less prominently, FC continues to be featured in a positive light within some prominent media outlets (e.g., *Time* magazine; *www.time.com/time/nation/article/0,8599,1192775,00.html*). Moreover, despite a continued lack of evidence for the procedure, many parents and special education teachers still rely on FC as a means of communication for individuals with ASD (Hall & Riccio, 2012; Hess, Morrier, Heflin, & Ivey, 2008).

In his seminal article on FC, "Communication Unbound: Autism and Apraxia," Biklen (1990a) described how Crossley (a pioneer in FC) first applied FC to individuals with cerebral palsy in Australia to guide their hand/finger movements when using a picture board to communicate. Crossley is credited with introducing FC to children with autism, although Biklen popularized the procedure in the United States and many other countries. FC has since been extended for use with populations with an array of communication disorders.

Advocates present FC as a means of facilitating language expression for individuals with absent or limited communicative abilities and assert that by assisting individuals to control their motor movements, hidden language abilities will emerge. To improve communication skills in children, a variety of communicative devices are used, including electronic typewriters, handheld computers, keyboards, and letter and picture boards (Biklen, 1990a). The following is a description of FC in a memo from Biklen for teachers, speech pathologists, and others interested in autism and FC:

> The method involves initial hand-over-hand and/or arm support, pulling the hand back after each selection, slowing down the movements, assistance in isolating the index finger, verbal reassurances, and encouragement. Over time, the physical support can be faded back completely or to just a hand on the shoulder. (1990b, p. 1)

Similarly, "supported typing" has more recently been described by the Institute of Communication and Inclusion at Syracuse University as a method that

> involves a communication partner who may provide emotional encouragement, communication supports (e.g., monitoring to make sure the person looks at the keyboard and checks for typographical errors) and a variety of physical supports, for example to slow and stabilize the person's movement, to inhibit impulsive pointing, or to spur the person to initiate pointing; the facilitator should never move or lead the person. (*http://soe.syr.edu/centers_institutes/institute_communication_inclusion/what_is_supported*_typing/default.aspx)[1]

Soon after FC began to rise in popularity, several scientists evaluated its validity. Specifically, Jacobson and Mulick (1992) noted that the demonstration of FC requires evidence that the individual is communicating independently, and not being influenced by the facilitator. One method used to remove potential influence from the facilitator is to pose questions to the affected individual, but not the facilitator (Dayan & Minnes, 1995). Results have shown that when facilitators are unable to hear the questions, or when they hear conflicting information, the individual consistently responds incorrectly (Bebko, Perry, & Bryson, 1996). Similarly, Wheeler, Jacobson, Paglieri, and Schwartz (1993) assessed differences in performance when the facilitator was able to see or not see what was being presented. This study used a table with a divider in the middle that permitted the facilitator and child to see different views, while allowing the facilitator to provide FC for the child. The results of this study revealed no support for FC and suggested a strong facilitator influence on the children's responses. Another ("message-passing") methodology presents the individual with questions or information in the absence of a facilitator. Later, the facilitator joins the individual to answer questions about the events prior to the presence of the facilitator. Using this method, Regal, Rooney, and Wandas (1994) reported that no correct responses were made, even though the facilitator was confident that the individual performed well.

Smith, Haas, and Belcher (1994) conducted a systematic investigation to examine the effects of facilitator knowledge and level of assistance

[1]It is important to distinguish augmentative communication from FC. Augmentative communication is an empirically demonstrated method of aiding a person with communication, sometimes using devices similar to those used in FC. However, in augmentative communication, devices are modified to allow direct voluntary control, using even the most subtle muscle movements. This permits independent communication (Cummins & Prior, 1992). In contrast, FC or supported typing uses the facilitator as an "intermediary" for communication purposes. Thus, what distinguished FC from augmentative communication is the procedure, not the devices and labels given to the procedure.

with amount of facilitator influence. In half of the trials, the facilitator was aware of the stimulus presented to the child. Three levels of facilitator support were examined: no help, hand-over-hand assistance without preventing errors, and hand-over-hand with preventing errors. The authors found that correct responses occurred only when the facilitator was aware of the stimulus and full support was provided.

To further evaluate the influence of physical support, Edelson, Rimland, Berger, and Billings (1998) provided a hand-support device. Individuals with autism were taught to type using three methods: a human-facilitated condition, a mechanical-facilitated condition, and a nonfacilitated condition (i.e., the individual received no form of assistance when typing). After 8 weeks, a nonfacilitated maintenance test of performance showed no evidence of learning in any of the facilitated conditions.

Proponents of FC claimed the experimental studies conducted were inappropriately designed and did not accurately measure performance. Silliman (1995) asserted that the studies were conducted out of the participants' normal social context, creating an unfamiliar environment that hindered performance. Duchan (1995) stated that "the context of interaction is not a naturally occurring one, but one that is tampered with in a variety of ways" (p. 208).

To address such criticisms, Kerrin, Murdock, Sharpton, and Jones (1998) conducted a well-designed study in children's typical classroom settings, thereby addressing the concern regarding unfamiliar environments. The classroom teacher and speech pathologist conducted the study. The two subjects in this study were diagnosed with autism and were familiar with both instructors. For one week prior to the study, the speech pathologist wore sunglasses. When the study began, she continued to wear the sunglasses in the classroom and while facilitating with both subjects. Throughout the day, children received a picture/written word labeling task that required them to point to the correct picture or word after a verbal request (e.g., "Point to . . . "). The two facilitator conditions, blind (i.e., the facilitator wore sunglasses with cardboard on the lenses to prevent sight) or sighted (e.g., the facilitator wore sunglasses only, enabling sight), were conducted on alternating days. The results showed that the subjects' responses were influenced significantly by the facilitator's ability to see the stimuli. The facilitator reported she did not believe she was intentionally influencing the subjects' responses.

Emerson, Grayson, and Griffiths (2001) conducted controlled analyses of FC and analyzed FC transcripts to assess the validity of FC among 14 participants. Importantly, proponents of FC conducted this study, so that the sessions were presumably designed to address the FC community's criticisms of previously published studies (e.g., the lack of testing in a familiar environment). The authors found no evidence for FC using the controlled experimental tasks. They suggested that the transcripts seemingly supported the efficacy of FC in that the FC users "self-disclosed" information

that facilitators were unlikely to guess. Nevertheless, these findings provide only circumstantial and unconvincing evidence for FC.

Grayson, Emerson, Howard-Jones, and O'Neil (2012) conducted the only eye-tracking study of FC. The authors examined the gaze patterns of an individual with ASD who had been using FC for over 5 years. For approximately two hours, an eye-tracker recorded the individual's gaze movements while a facilitator assisted the individual in typing on an augmentative device. The authors hypothesized that authorship could be attributed to the individual if fixations on to-be-typed letters were on average longer than fixations on not-to-be typed letters. They also suggested that a systematic relationship between eye gaze movements and pointing was indicative of client authorship. As predicted by the authors, the individual fixated on letters to-be-typed significantly more than other letters. Additionally, the place where the participant looked was correlated significantly to the places at which he pointed. Several flaws exist in the logic of this design. Importantly, the fact that the individual's gaze followed his arm movements does not address the issue of client authorship. As the authors note, a plausible alternative explanation is that the pointing causes the relationship between looking and pointing, with pointing attributed to the facilitator.

Mostert (2001, 2010) conducted two thorough reviews of the FC literature. Across both reviews, studies with adequate control procedures and methodological rigor revealed no evidence for FC, whereas studies with poor control procedures and methodological problems were more likely to claim support for FC. Earlier reviews of FC reached similar conclusions, suggesting that there is no compelling evidence for FC (Cummins & Prior, 1992; Green, 1994; Hudson, 1995; Jacobson, Mulick, & Schwartz, 1995; Simpson & Myles, 1995).

As Mostert (2010) noted, research on FC waned significantly following 2001. In the past 12 years, only one carefully controlled experimental study has been conducted on FC (Wegner, Fuller, & Sparrow, 2003). Thus, proponents of FC and individuals who work within the field of FC conduct the majority of recently published studies on the topic (Mostert, 2010). Given this discrepancy and the lack of serious empirical interest in FC, "It is probable that pro-FC publications will proliferate, leading the field once again to another recurrent saga of the ineffective masquerading as the effective and wishful thinking as incontrovertible evidence" (Mostert, 2010, p. 39).

Supporters of FC advise that the "facilitator should never move or lead the person" (*http://soe.syr.edu/centers_institutes/institute_communication_inclusion/what_is_supported_typing/default.aspx*). Indeed, FC facilitators are even encouraged to be "looking for ways to make sure they [we] are not influencing the child" (Boynton, 2012, p. 5). Wegner et al. (2003) conducted a series of experiments to examine the ability of facilitators to refrain from influencing communication when they know the information

being asked of the client. "Facilitators" (undergraduate university students) were instructed to "read the muscle movements" of a confederate, who played the role of a client. The confederates wore headphones and were thus unable to hear any questions asked (unbeknownst to the facilitators). Facilitators reported they were not responsible for client responses; however, the data clearly identified a response pattern that required facilitator interference. Wegner et al. (2003) also found that facilitators who believed strongly in FC were more likely to interpret the confederate's responses as genuine. Thus, it may be difficult for facilitators to recognize their influence on the communicative responses of clients. A strong belief in FC and a genuine interest in assisting individuals with developmental disabilities may further cloud the judgment of FC facilitators (for a personal account of this phenomenon from the perspective of a former facilitator, see Boynton, 2012).

Given the complete lack of any supportive evidence for FC, several ethical implications arise concerning the use of FC in applied settings. One concerns the determination of educational placement. Because FC provides false and misleading evaluations of the individual's intellectual and achievement level, incorrect educational placement can result. Children communicating with FC have been placed in grades inconsistent with their ability level (Beck & Pirovano, 1996; Jacobson & Mulick, 1992), an error that can promote false hopes and expectations of parents and facilitators, as well as denial of appropriate educational services.

Another important ethical issue concerning FC is its use in determining allegations of sexual and physical abuse by a parent or relative. Siegel (1995) assessed allegations of sexual molestation of two adolescents made through FC. In an assessment with two trained facilitators who were unaware of the allegations, both adolescents showed random responses to both open-ended questions and objective information. Neither adolescent showed negative affect toward their fathers. Unfortunately, despite court precedence in these early cases, sexual allegations made using FC were used as recently as 2007 in Michigan to prosecute the father of a teenage daughter with autism for sexual abuse charges (Gomstyn, 2012). The charges were eventually dropped, and the family received financial compensation for civil damages from the police department involved in the case.

Due to the lack of scientific evidence for the treatment and the potential for harm for individuals and their families, the American Psychological Association adopted the position that "facilitated communication is a controversial and unproved communication procedure with no scientifically demonstrated support for its efficacy" (*www.apa.org/divisions/div33/fcpolicy.html*). Similar policies have been adopted by the American Academy of Child and Adolescent Psychiatry, the American Association of Mental Retardation, and the American Academy of Pediatrics (Science and Pseudoscience Review in Mental Health, 1992).

Dolphin-Assisted Therapy

As with many other pseudoscientific treatments, dolphin-assisted therapy (DAT) was widely popularized by the media. In 1998, DAT was presented on Cable News Network as a procedure in which a child completes a one-to-one teaching session and then earns the reward of swimming with a dolphin. The child's success in the one-to-one teaching session was attributed to the motivating effects of swimming with a dolphin. Of note, dolphins are currently the only nondomesticated animals utilized as treatment partners for children with autism. Proponents of dolphin-assisted therapy argue that the technique assists in treating not only ASD, but a variety of other psychological and biological disorders, including Down syndrome, depression, blindness, AIDS, cancer, and enuresis (Wermer, 2008; Batozsky, 2012). An estimated 100 organizations offer DAT (Wermer, 2008).

There appear to be no set criteria for what constitutes DAT, as there is no regulation of the purportedly therapeutic procedure. Thus, the term "DAT" encompasses a variety of interactions with dolphins aimed to alleviate psychological or medical symptoms. Typically, DAT involves individuals with autism interacting with dolphins at the poolside as a reward for completing a task. Rewards may include petting the dolphin, feeding the dolphin, and dorsal fin rides (Williamson, 2008). The typical charge for dolphin therapy is $2,600 per week (or five, 40-minute sessions) (Brakes & Williamson, 2007). This cost excludes airfare and lodging, which add an additional cost to families pursuing the treatment. The significant cost and time required for this treatment may foster an expectation of positive results.

Some proponents of DAT have suggested that the technique works through sonophoresis—"The increasing of enzymes' and special hormones' stream, penetrating through cells' membranes as the result of cavitation under the influence of ultrasound" (*www.dolphintherapy.ru/en/sonoforez.shtml*). Thus, the dolphins' ultrasonic echolocation influence is thought to cause "positive chemical and electrical changes" in the individual. However, the claim that echolocation serves as the therapeutic agent in DAT has no scientific support. Furthermore, systematic empirical observations of dolphins and their interactions with individuals with disabilities have led to the conclusion that dolphins fail to independently aim their head towards the patient for any significant amount of time, suggesting that ultrasound effects are not possible (Brensing & Linke, 2003).

Nathanson (1998) conducted an evaluation of children who underwent a 1- to 2-week program of DAT. He reported that the treatment provided long-lasting effects across several areas of impairment, including speech, gross and fine motor skills, and attention. Nevertheless, a methodological analysis of his report by Marino and Lilienfeld (1998) reveals serious methodological flaws in this investigation. Nathanson's research contained several threats to internal validity, including (1) the lack of a

control group, rendering it impossible to determine whether the effects found were solely due to DAT rather than other factors and (2) a failure to address maturation as an explanation for changes in participant characteristics. In addition, Nathanson's questionnaire assessing parents' subjective reports of their child's change over time was vulnerable to demand characteristics and expectation bias. For example, Nathanson acknowledged that, "each behavioral item was preceded by the statement, 'As a result of Dolphin Human Therapy, my child has maintained or improved his/her ability to. . . .' " Moreover, Nathanson's conclusions are difficult to evaluate independently because the data reported are overall means for combined samples rather than individual statistics.

In a subsequent review of DAT, Marino and Lilienfeld (2007) examined five peer-reviewed DAT studies. Two studies examined included participants with an ASD (i.e., Lukina, 1999; Servais, 1999). Marino and Lilienfeld identified several methodological flaws in both studies, including (1) expectancy and demand effects; (2) novelty effects (improvement resulting from the administration of a new and interesting intervention); and (3) the use of subjective and non-standardized instruments to measure change. The authors concluded that "Despite DAT's extensive promotion to the general public, the evidence that it produces enduring improvements in the core symptoms of any psychological disorder is nil" (Marino & Lilienfeld, 2007, p. 248).

Despite the lack of scientific evidence for DAT, the treatment remains popular. Furthermore, new forms of DAT have evolved in recent years, including the use of robotic dolphins and virtual dolphin therapy. Nathanson (2007) suggested that robotic dolphins are also effective as therapeutic "partners" for children with developmental disabilities. Although the author suggested that the intervention is more cost-effective than live dolphin treatment, the basis for this conclusion is unclear, as trained staff members are needed to control the robot dolphin. Virtual dolphin therapy (*http://virtualdolphintherapy.com*) was also recently highlighted on the CBS show *The Doctors* as a "multisensory" experience that can produce relaxing and calming effects. Nigel Collier, a father of a child with a developmental disability and a marine engineer, was also recently featured in BBC News for his invention termed the "Dolphin Dome" (*www.bbc.co.uk/news/uk-england-humber-13544115*). The device uses video screens and sounds of the sea to recreate the experience of being in the ocean with dolphins. No scientific investigations of the effectiveness of virtual dolphin therapy for children with ASD currently exist.

The Whale and Dolphin Conservation Society has spoken publicly against the practice of DAT and highlighted the damage incurred by dolphins who suffer from confinement and human disturbance as a result of the therapy. To date, no independent, well-controlled studies support claims of DAT's effectiveness and the claim that dolphin echolocation produces beneficial cellular metabolic changes in individuals with ASD.

Chiropractic Manipulation

A relatively new and increasingly popular alternative treatment for ASD is chiropractic manipulation. As of this writing, a Google search of the terms "autism" and "chiropractic" yielded 1,490,000 results. Comparatively, a similar search for "autism" and "dolphin therapy" yielded only 16,300 results, and a search for "autism" and "facilitated communication" yielded only 60,400 hits. Moreover, an estimated 10% of parents of children with ASD have used chiropractic care for their children in the hopes of alleviating ASD-related symptoms (Hanson et al., 2007; Wong & Smith, 2006). The popularity of this alternative treatment is borne out by the development of treatment guidelines for the chiropractic care of children with ASD and other special needs (Barnes, 1997; Jennings & Barker, 2004).

Proponents of chiropractic care suggest that cranial misalignments contribute to ASD symptoms. "When the skull is misaligned, the part of the brain adjacent to the misalignment will suffer a greater pressure, a decreased blood and nerve supply and result in hypoxia (lack of oxygen) to that part of the brain. This adversely affects how that part of the brain functions and can result in Autistic characteristics" (*www.turnerwellness.com/pdf/Autism_Article_2010.pdf*). Similarly, Alcantara, Alcantara, and Alcantara (2011) argued that chiropractic care of children with ASD works via "some type of correction or mitigation of abnormal sensory information" (Alcantara et al., 2011, p. 388), although the authors contended that the "exact mechanisms through which chiropractic may be helpful in the treatment of ASD is not known" (Alcantara et al., 2011, p. 388). Purported benefits of chiropractic sessions vary, although improvements in language, social skills, problematic behaviors (repetitive behaviors, aggression, hyperactivity), and dietary difficulties (restrictiveness and gastrointestinal problems) have all been suggested (Aguilar, Grostic, & Pfleger, 2000; Hoffmann & Russell, 2008; Jennings & Baker, 2004; Khorshid, Sweat, Zemba, & Zemba, 2006; Marni & Marni, 2010; McCormick, 2008). The number of clinical sessions for the treatment has yet to be established, although across published studies, chiropractic visits range from 3 visits per week for 10 weeks and 1–2 visits per month for 9 months (Alcantara et al., 2011).

The empirical evidence in support of chiropractic care for individuals with ASD is lacking. Of the handful of peer-reviewed articles on the subject, most are case and observational studies (Alcantara et al., 2011; Ferrance, 2003; Jennings & Barker, 2004). Alcantara et al. (2011) identified a total of five peer-reviewed articles consisting of three case reports, one cohort study, and one randomized *comparison* trial. The three case reports (Hoffmann & Russell, 2008; Marni & Marni, 2010; McCormick, 2008) are plagued with methodological problems, including (1) exclusive reliance on parent report for evidence of improvements in ASD symptoms, (2) continuation of other treatments during chiropractic care, and (3) lack of control for maturation. Moreover, one case study (Hoffmann & Russell,

2008) relied on subjective parent report of improvements and neglected to use a pre–post questionnaire to assess improvement. In the cohort study, Aguilar et al. (2000) administered chiropractic adjustments to 26 children with ASD over 9 months. The authors assessed behavioral improvements using standardized instruments (the Modified Autism Checklist; Robins, Fein, Barton, & Green, 2001 and the Childhood Autism Rating Scale; Schopler, Reichler, & Renner, 1988) and reported positive changes across pre- and posttreatment ratings. However, the lack of a control group seriously compromises the interpretation of the results. In the only other group design, Khorshid et al. (2006) compared the effectiveness of two types of chiropractic adjustments for autism. The authors used standardized instruments developed to monitor progress in ASD and concluded that one chiropractic treatment (full spine spinal manipulative therapy) was associated with more improvement than the other treatment (atlas orthogonal upper cervical spinal manipulative therapy). Nevertheless, the authors did not conduct statistical analyses, relying instead on mean score differences to draw conclusions, a particularly problematic practice given the low sample size of each group (n = 7). The lack of a true control group (participants who did not receive chiropractic care) is also problematic, as maturation effects are not accounted for. Lastly, as both groups received chiropractic care, expectancy effects may explain some of the improvement in parental report scores.

In sum, no rigorous empirical evaluation of chiropractic care for individuals with ASD exists. Moreover, the mechanism by which chiropractic manipulation is purported to alleviate ASD symptoms is unclear and scientifically unsubstantiated.

This section examined only a handful of pseudoscientific treatments marketed for ASD. Over 400 treatments have been purported to ameliorate ASD-related symptoms, and only a small minority show good empirical support (Romanczyk et al., 2008). Thus, identifying treatments with empirical support is vitally important, as evidenced by a plethora of treatment review articles and documents (for the most recent large-scale review, see National Autism Center, 2009). Based on recent reviews, the following sampling of treatments is frequently marked by weak or absent empirical support:

- Auditory integration training (Herbert, Sharp, & Gaudiano, 2008; Schechtman, 2007; Zimmer, 2011; Tuzikow & Holburn, 2011).
- Chelation therapy (Tuzikow & Holburn, 2011; McDonald, Pace, Blue, & Schwartz, 2012; Zimmer, 2011).
- Gluten- and casein-free diets (Herbert, Sharp, & Gaudiano, 2008; Schechtman, 2007).
- Hyperbaric oxygen therapy (Schechtman, 2007; Tuzikow & Holburn, 2011).
- Sensory integration training (Herbert, Sharp, & Gaudiano, 2008).

ANALYSIS OF PARENT AND PROFESSIONAL PREFERENCES

Reviews consistently show that families prefer interventions for ASD that lack an evidential base (Green et al., 2005; Romanczyk et al., 2008). Interestingly, many providers of evidence-based interventions also use nonevidence-based and pseudoscientific treatments (Schrek & Mazur, 2008). In the context of limited resources and an increasing prevalence of questionable interventions, this situation is unfortunate and raises the question of why these treatments remain popular?

Mackintosh, Goin-Kochel, and Myers (2012) surveyed a large sample (N = 486) of parents regarding what they liked and disliked about treatments. Parent likes and dislikes were unrelated to the type of interventions; however, their ratings were highly related to respondents' interpersonal interactions with service providers. Although systematic evidence is limited, our anecdotal experience suggests that parents' "first contact" with a service provider is highly influential in treatment selection.

In an earlier review (Romanczyk & Gillis, 2004), we summarized our clinical experience regarding the question of how caregivers approach treatment choice decisions. We observed seven nonmutually exclusive strategies:

1. "They know what's best"—Place trust in a service provider.
2. "Hedge your bets"—Do a little bit of everything.
3. "Fanatical focus"—Pursue a single course with overwhelming intensity and focus.
4. "Hope for the best"—Forgo formal treatment and participate in typical activities that are available.
5. "Cure du jour"—Pursue new treatments as they appear and drop the current program.
6. "A friend told me"—Do what seemed to work for the child of someone you know or have read about.
7. "Guru selection"—Follow and believe in a single, specific "expert."

These strategies are all understandable given information overload, disagreements among service providers regarding effective treatments, and the unique stressors faced by parents of children with an ASD. Navigating the complex world of service delivery, exploring information in the context of all of the demands family and children present, and synthesizing vast amounts of information and misinformation (e.g., a Google search of "autism treatment" yields 47,300,000 results) is a daunting task. Hence, it is not surprising that methodologically sound practice guidelines have only begun to staunch the flow of pseudoscientific practices for ASD.

On a more positive note, throughout the last decade, many parents, caregivers, and service providers have partnered with Autism Speaks (*www.autismspeaks.org*), the nation's largest autism science and advocacy organization, to reform state and federal health care laws regulating the

funding of IBI (a.k.a. ABA) for ASD. Autism Speaks has been instrumental in this effort, as its mission is to promote the implementation and dissemination of evidence-based practice in the treatment of ASD. Another one of its major goals is to create legislation that requires insurance companies to provide coverage for evidence-based behavioral health services, namely IBI, for ASD.

Prior to this insurance reform initiative, health care plans rarely covered IBI. Between 2001 and 2006, only one state had enacted a law requiring state-regulated insurance plans to cover behavioral health services for ASD. However, since 2006, the support for this initiative has gained substantial momentum (see Figure 14.1), with an average of 4.5 states per year adopting autism insurance mandates. By 2013, there were 34 states, plus the District of Columbia, with an autism insurance mandate, and there have already been 2 states in 2014 to enact similar laws (Autism Speaks, 2014).

Given that these laws and regulations are being passed at the state level, insurance mandates across states differ in significant details. Most states differ in regard to the dollar or age capitation (cap) placed on services, how IBI is defined, and in what settings IBI is covered (e.g., small group vs. individual). In many states, once a law is passed, reform is still needed to allow for adequate coverage. For example, some states provide coverage only for younger children (e.g., under 6 years old), excluding the large population of school-age children, adolescents, and adults with ASD in need of IBI. Despite the potentially long road ahead with regard to refining the autism

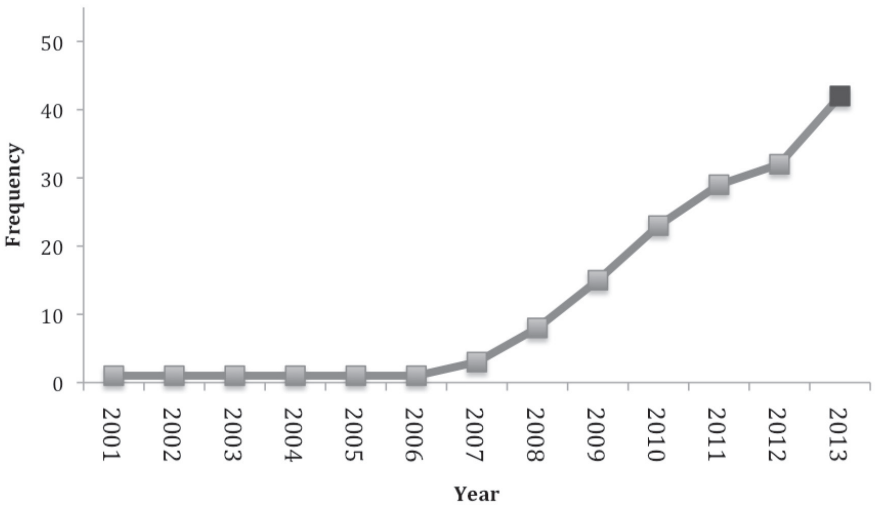

FIGURE 14.1. The number of states by year that passed an autism insurance mandate. The District of Columbia is also included.

mandates, these advances are encouraging and hold out the promise that more individuals with ASD will receive scientifically based interventions.

GLOSSARY

Applied behavior analysis (ABA): The scientific study of behavior that uses a specific conceptual and methodological approach for direct quantification and analysis of behavior. It historically derives primarily from research in psychology on basic learning processes. Applied behavior analysis refers to the precise measurement and analysis of behavior and learning patterns and the conditions that serve to elicit and maintain these behaviors and patterns.

Auditory integration training: Repeated presentation of modified sounds hypothesized to alleviate atypical responses to sound in children with learning and developmental disabilities.

Elimination diet: Any eating plan that eliminates a specific food group from the diet. In autism, foods commonly eliminated include those containing gluten (found in wheat products) or casein (found in milk products).

Facilitated communication (FC): A technique that involves providing physical assistance with the use of a communicative device (e.g., computer keyboard, letter board) to aid persons with autism to control motor movements to permit nonverbal communication.

Floortime: Child-directed playtime used to teach the six fundamental milestones that are the basis of Stanley Greenspan's developmental-individual difference, relationship-based model.

Secretin: A polypeptide hormone that causes the pancreas to secrete digestive enzymes that help to process food in the small intestines.

Sensory integration therapy: Client-directed therapy whose stated goal is to enable the child to better process and organize information from the sensory world using a variety of equipment designed to stimulate various sensory systems (e.g., balls, scooter boards, nets).

Splinter skills: Markedly above-average skills in one domain, a phenomenon observed in approximately 10% of children with infantile autism.

Vitamin B_6 treatment: A vitamin (B_6) that is typically administered in combination with magnesium and is presumed to reduce problem behaviors in individuals with autism.

REFERENCES

Aguilar, A. L., Grostic, J. G., & Pfleger, B. (2000). Chiropractic care and behaviour in autistic children. *Journal of Clinical Chiropractic Pediatrics*, *1*, 293–304.

Alcantara, J., Alcantara, J. D., & Alcantara, J. (2011). A systematic review of the literature on the chiropractic care of patients with autism spectrum disorder. *Explore*, *7*(6), 384–390.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

American Psychological Association. (1994, August 14). *Resolution on facilitated communication by the American Psychological Association.* Available at *www.apa.org/divisions/div33/fcpolicy.html.*

Barnes, T. (1997). Chiropractic management of the special needs child. *Topics in Clinical Chiropractic, 4*(4), 9–18.

Batozsky, J. (2012). *Sonophoretic model in dolphin assisted therapy.* Available at *www.dolphintherapy.ru/en/sonoforez.shtml.*

Bebko, J. M., Perry, A., & Bryson, S. (1996). Multiple method validation study of facilitated communication: II. Individual differences and subgroup results. *Journal of Autism and Developmental Disorders*, *26*(1), 19–42.

Beck, A. R., & Pirovano, C. M. (1996). Facilitated communicators' performance on a task of receptive language. *Journal of Autism and Developmental Disorders*, *26*(5), 497–512.

Bertoglio, K., James, S. J., Deprey, L., Brule, N., & Hendren, R. L. (2010). Pilot study of the effect of methyl b12 treatment on behavioral and biomarker measures in children with autism. *Journal of Alternative and Complementary Medicine, 16, 555–560.*

Biklen, D. (1990a). Communication unbound: Autism and praxis. *Harvard Educational Review*, *60*(3), 291–314.

Biklen, D. (1990b). *Information packet on facilitated communication.* Syracuse, NY: Syracuse University.

Boynton, J. (2012). Facilitated communication—what harm it can do: Confessions of a former facilitator. *Evidence-Based Communication Assessment and Intervention*, *6*(1), 3–13.

Bradley, E. A., Summers, J. A., Wood, H. L., & Bryson, S. E., (2004). Comparing rates of psychiatric and behavior disorders in adolescents and young adults with severe intellectual disability with and without autism. *Journal of Autism and Developmental Disorders*, *34*(2), 151–161.

Brensing, K., & Linke, K. (2003). Behavior of dolphins towards adults and children during swim-with-dolphin programs and towards children with disabilities during therapy sessions. *Anthrozoös*, *16*(4), 315–331.

Centers for Disease Control and Prevention. (2012, March 30). Morbidity and Mortality Weekly Report. In *Surveillance Summaries*, *61*(6), 1–19.

Centers for Disease Control and Prevention. (2013, March 20). *National Health Statistics Reports, Number 65.*

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 17–18.

Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *Clinical Psychologist, 49*, 5–18.

Christon, L. M., Mackintosh, B. J., & Myers, B. J. (2010). Use of complementary and alternative medicine (CAM) treatments by parents of children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 4,* 249–259.

Croen, L. A., Grether, J. K., Hoogstrate, J., & Selvin, S. (2002). The changing

prevalence of autism in California. *Journal of Autism and Developmental Disorders*, *32*(3), 207–215.

Cummins, R. A., & Prior, M. P. (1992). Autism and assisted communication: A response to Biklen. *Harvard Educational Review*, *62*(2), 228–241

Dawson, G. (2012, June 11). DSM-5 update: Our letter to the revision committee. *Autism speaks*. Retrieved from *www.autismspeaks.org/blog/2012/06/11/dsm-5-update-our-letter-revision-committee?&utm_source=social-media&utm_medium=text-link&utm_campaign=espeaks*.

Dayan, J., & Minnes, P. (1995). Ethical issues related to the use of facilitated communication techniques with persons with autism. *Canadian Psychology, 36*(3), 183–189.

DeStefano, F., Price, C. S., & Weintraub, E. S. (2013, April). Increasing exposure to antibody-stimulating proteins and polysaccharides in vaccines is not associated with risk of autism. *Journal of Pediatrics*, *163*(2), 561–567.

Duchan, J. F. (1995). The role of experimental research in validating facilitated communication: A reply. *Journal of Speech and Hearing Research*, *38*, 206–210.

East Yorkshire man invents dolphin therapy device. (2011, May 25). Available at *www.bbc.co.uk/news/uk-england-humber-13544115*.

Edelson, S. M., Rimland, B., Berger, C., & Billings, D. (1998). Evaluation of a mechanical hand-support for facilitated communication. *Journal of Autism and Developmental Disorders*, *28*(2), 153–157.

Eikeseth, S. (2009). Outcome of comprehensive psycho-educational interventions for young children with autism. *Research in Developmental Disabilities, 30*, 158–178.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child and Adolescent Psychology, 38*, 439–450.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., & Eikeseth, S. (2010). Using participant data to extend the evidence base for intensive behavioral intervention for children with autism. *American Association on Intellectual and Developmental Disabilities, 115*(5), 381–405.

Emerson, A., Grayson, A., & Griffiths, A. (2001). Can't or won't?: Evidence relating to authorship in facilitated communication. *International Journal of Language and Communication Disorders*, *36*, 98–103.

Ferrance, R. J. (2003). Autism—another topic often lacking facts when discussed within the chiropractic profession. *Journal of the Canadian Chiropractic Association*, *47*(1), 4.

Field, M. J., & Lohr, K. N. (1992). *Guidelines for clinical practice: From development to use. Institute of Medicine*. Washington, DC: National Academy Press.

Findling, R. L., Maxwell, K., Scotese-Wojtila, L., Huang, J., Yamashita, T., & Wiznitzer, M. (1997). High-dose pyridoxine and magnesium administration in children with autistic disorder: An absence of salutary effects in a double-blind, placebo-controlled study. *Journal of Autism and Developmental Disorders*, *27*(4), 467–478.

Ghaziuddin, M., Weidmer-Mikhail, E., & Ghaziuddin, N. (1998). Comorbidity of Asperger syndrome: A preliminary report. *Journal of Intellectual Disability Research*, *42*(4), 279–283.

Golnik, A. E., & Ireland, M. (2009). Complementary alternative medicine for children with autism: A physician survey. *Journal of Autism and Developmental Disorders*, *39*(7), 995–1005.

Gomstyn, A. (2012, January 7). *Not just the wendrows: Sex abuse cases dismissed after faciliatated communication*. Available at *http://abcnews.go.com/Health/wendrows-sex-abuse-cases-dismissed-facilitated-communication/story?id=15274276*.

Grayson, A., Emerson, A., Howard-Jones, P., & O'Neil, L. (2012). Hidden communicative competence: Case study evidence using eye-tracking and video analysis. *Autism*, *16*(1), 75–86.

Green, G. (1994). Facilitated communication: Mental miracle or sleight of hand? *Behavior and Social Issues*, *4*(1–2), 69–85.

Green, V. A., Pituch, K. A., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2005). Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities*, *27*, 70–84.

Greenspan, S. I., Prizant, B. M., Wetherby, A. M., & First Signs. (2001). *Is your baby meeting these important milestones?: Key social, emotional, communication, and behavior milestones for your baby's healthy development*. Merrimac, MA: First Signs.

Greenspan, S. I., & Shanker, S. I. (2004). *The first idea*. Cambridge, MA: Da Capo Press.

Greenspan, S. I., & Wieder, S. (1998). *The child with special needs: Encouraging intellectual and emotional growth*. Reading, MA: Perseus Books.

Greenspan, S. I., & Wieder, S. (1999). A functional developmental approach to autism spectrum disorders. *Journal of the Association for Persons with Severe Handicaps, 24*(3), 147–161.

Greenspan, S., & Wieder, S. (2000). The assessment and diagnosis of infant disorders: Developmental level, individual differences, and relationship-based interactions. *Early Intervention, Evaluation, and Assessment I*, *2*, 207–237.

Greenspan, S. I., DeGangi, G., & Wieder, S. (2001). *The functional emotional assessment scale (FEAS)*. Bethesda, MD: Interdisciplinary Council on Developmental and Learning Disorders.

Hall, S. E., & Riccio, C. A. (2012). Complementary and alternative treatment use for autism spectrum disorders. *Complementary Therapies in Clinical Practice*, *18*(3), 159–163.

Hanson, E., Kalish, L. A., Bunce, E., Curtis, C., McDaniel, S., Ware, J., et al. (2007). Use of complementary and alternative medicine among children diagnosed with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*, 628–636.

Herbert, J. D., Sharp, I. R., & Gaudiano, B. A. (2002). Separating fact from fiction in the etiology and treatment of autism: A scientific review of the evidence. *Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work, 1*(1), 23–43.

Herbert, J., Sharp, I. R., & Gaudino, B. A. (2008). Separating fact from fiction in the etiology and treatment of autism: A scientific review of the evidence. In S. O. Lilienfeld, J. Ruscio, & S. J. Lynn (Eds.), *Navigating the mindfield: A guide to separating science from pseudoscience in mental health* (pp. 187–237). Amherst, NY: Prometheus Books.

Hess, K. L., Morrier, M. J., Heflin, L., & Ivey, M. L. (2008). Autism treatment survey: Services received by children with autism spectrum disorders in public school classrooms. *Journal of Autism and Developmental Disorders*, *38*(5), 961–971.

Hilton, J. C., & Seal, B. C. (2007). Brief report: Comparative ABA and DIR trials in twin brothers with autism. *Journal of Autism and Developmental Disorders*, *37*(6), 1197–1201.

Hoffmann, N., & Russell, D. (2008). Improvement in a 3½-year-old autistic child following chiropractic intervention to reduce vertebral subluxation. *Journal of Vertebral Subluxation Research,* 1–4.

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, *26*(4), 359–383.

Hudson, A. (1995). Disability and facilitated communication: A critique. *Advances in Clinical Child Psychology; Advances in Clinical Child Psychology*, *17*, 197–232.

Jacobson, J. W., & Mulick, J. A. (1992). Speak for yourself, or . . . I can't quite put my finger on it! *Psychology in Mental Retardation and Developmental Disabilities*, *17*(3), 37.

Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1995). A history of facilitated communication: Science, pseudoscience, and antiscience science working group on facilitated communication. *American Psychologist*, *50*(9), 750–765.

Jennings, J., & Baker, M. (2004). Autism: A chiropractic perspective. *Clinical Chiropractic*, *9*, 6–10.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kaplan, R. (Producer). (1992, February). Free from silence. In *Prime Time Live*. New York: ABC News.

Kerrin, R. G., Murdock, J. Y., Sharpton, W. R., & Jones, N. (1998). Who's doing the pointing?: Investigating facilitated communication in a classroom setting with students with autism. *Focus on Autism and Other Developmental Disabilities*, *13*(2), 73–79.

Khorshid, K. A., Sweat, R. W., Zemba, D. A., & Zemba, B. N. (2006). Clinical efficacy of upper cervical versus full spine chiropractic care on children with autism: A randomized clinical trial. *Journal of Vertebral Subluxation Research*, 1–7.

King, M., & Bearman, P. (2009). Diagnostic change and the increased prevalence of autism. *International Journal of Epidemiology*, *38*, 1224–1234.

Kozlowski, B. W. (1992). Megavitamin treatment of mental retardation in children: A review of effects on behavior and cognition. *Journal of Child and Adolescent Psychopharmacology, 2*(4), 307–320.

Kuriyama, S., Kamiyama, M., Watanabe, M., Tamahashi, S., Muraguchi, I., Watanabe, T., et al. (2002). Pyridoxine treatment in a subgroup of children with pervasive developmental disorders. *Developmental Medicine and Child Neurology*, *44*(4), 284–286.

Lord, C., & Jones, R. M. (2012). Annual research review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 53*(5), 490–509.

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D., et al. (2011). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, *69*(3), 306–313.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, *55*, 3–9.

Lukina, L. N. (1999). Influence of dolphin-assisted therapy sessions on the functional state of children with psychoneurological symptoms of diseases. *Human Physiology*, *25*(6), 676–679.

Mackintosh, V. H., Goin-Kochel, R. P., & Myers, B. J. (2012). "What do you like/dislike about the treatments you're currently using?": A qualitative study of parents of children with autism spectrum disorders. *Focus on Autism and Developmental Disabilities*, *27*(1), 51–60.

Marino, L., & Lilienfeld, S. O. (1998). Dolphin-assisted therapy: Flawed data, flawed conclusions. *Anthrozoös*, *11*(4), 194–200.

Marino, L., & Lilienfeld, S. O. (2007). Dolphin-assisted therapy: More flawed data and more flawed conclusions. *Anthrozoös*, *20*(3), 239–249.

Marni, N., & Marni, S. (2010). Improvement in autism in a child coupled with reduction in vertebral subluxations: A case study and systematic review of the literature. *Journal of Pediatric, Maternal, and Family Health*, *3*, 107–115.

McCormick, J. (2008). Positive behavioral changes following chiropractic care in a child diagnosed with autism. *Journal of Clinical Chiropractic Pediatrics, 9*(1), 553–557.

McDonald, M. E., Pace, D., Blue, E., & Schwartz, D. (2012). Critical issues in causation and treatment of autism: Why fads continue to flourish. *Child and Family Behavior Therapy*, *34*(4), 290–304.

McEachin, J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation*, *97*(4), 359–372.

McPartland, J. D., Reichow, B., & Volkmar, F. R. (2012). Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(4), 368–383.

Missouri Department of Mental Health. (2012). Autism spectrum disorders: Guide to evidence-based interventions. Available at *www.autismguidelines.dmh.mo.gov*.

Mostert, M. P. (2001). Facilitated communication since 1995: A review of published studies. *Journal of Autism and Developmental Disorders*, *31*(3), 287–313.

Mostert, M. P. (2010). Facilitated communication and its legitimacy—twenty-first century developments. *Exceptionality*, *18*(1), 31–41.

Myers S., & Johnson, C. P. (2007). Management of children with autism spectrum disorders. *Pediatrics*, *120*(5), 1162–1182.

Nathanson, D. E. (1998). Long-term effectiveness of dolphin-assisted therapy for children with severe disabilities. *Anthrozoös*, *11*(1), 22–32.

Nathanson, D. E. (2007). Reinforcement effectiveness of animatronic and real dolphins. *Anthrozoös*, *20*(2), 181–194.

National Autism Center. (2009). *National Standards Report: National Standards Project—Addressing the need for evidence-based practice guidelines for autism spectrum disorders*. Randolph, MA: National Autism Center.

National Research Council. (2001). *Educating children with autism*. Washington, DC: National Academy Press.

New York State Department of Health, Early Intervention Program. (1999). *Clinical practice guideline: Guideline technical report. Autism/pervasive developmental disorders, assessment and intervention for young children (ages 0–3 years), No. 4217.* Albany: New York State Department of Health.

Noyes-Grosser, D. M., Holland, J. P., Lyons, D., Holland, C. L., Romanczyk, R. G., & Gillis, J. M. (2005). Rationale and methodology for developing guidelines for early intervention services for young children with developmental disabilities. *Infants and Young Children, 18*(2), 119–135.

Nye, C., & Brice, A. (2009). Combined vitamin B6-magnesium treatment in autism spectrum disorder. *Cochrane Collaboration.*

Pajareya, K., & Nopmaneejumruslers, K. (2011). A pilot randomized controlled trial of DIR/Floortime™ parent training intervention for pre-school children with autistic spectrum disorders. *Autism*, *15*(5), 563–577.

Palfreman, J. (Producer). (1993, October). Prisoners of silence. In *Frontline*. Alexandria, VA: Public Broadcasting Services.

Pauling, L. (1968, April). Orthomolecular psychiatry: Varying the concentrations of substances normally present in the human body may control mental disease. *Science*, *160*(3825), 265–271.

Pfeiffer, S. I., Norton, J., Nelson, L., & Schott, S. (1995). Efficacy of vitamin B6 and magnesium in the treatment of autism: A methodology review and summary of outcomes. *Journal of Autism and Developmental Disorders*, *25*(5), 481–493.

Regal, R. A., Rooney, J. R., & Wandas, T. (1994). Facilitated communication: An experimental evaluation. *Journal of Autism and Developmental Disorders*, *24*(3), 345–355.

Robins, D., Fein, D., Barton, M., & Green, J. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*, 131–151.

Romanczyk, R. G., Arnstein, L., Soorya, L., & Gillis, J. (2002). The myriad of controversial treatments for autism: A critical evaluation of efficacy. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in contemporary clinical psychology.* New York: Guilford Press.

Romanczyk, R. G., & Callahan, E. H. (2012). Autism and pervasive developmental disorders. In Vilanayur Ramachandran (Ed. in Chief), *The encyclopedia of human behavior.* Waltham, MA: Academic Press.

Romanczyk, R. G., & Gillis, J. M. (2004). Treatment approaches for autism: Evaluating options and making informed choices. In D. Zager (Ed.), *Autism: Identification, education and treatment* (3rd ed.). Hillsdale, NJ: Erlbaum.

Romanczyk, R. G., Gillis, J. M., White, S., & DiGennaro, F. (2008). Comprehensive treatment packages for ASD: Perceived vs. proven effectiveness. In J. L. Matson (Ed.), *Clinical assessment and intervention for autism spectrum disorders* (pp. 351–381). Burlington, MA: Elsevier.

Schechter, R., & Grether, J. K. (2008). Continuing increases in autism reported to California's developmental services system. *Archives in General Psychiatry, 65*(1), 19–24.

Schechtman, M. A. (2007). Scientifically unsupported therapies in the treatment

of young children with autism spectrum disorders. *Pediatric Annals, 36*(8), 497–505.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services.

Schrek, K. A., & Mazur, A. (2008). Behavior analyst use and beliefs in treatments for people with autism. *Behavioral Intervention, 23*, 201–212.

Science and Pseudoscience Review in Mental Health. (1992). Facilitated communication: Courts say "no." *Autism Research Review International*, *6*(3). Available at *www.srmhp.org/archives/facilitated-communication.html*.

Servais, V. (1999). Some comments on context embodiment in zootherapy: The case of the autidolfin project. *Anthrozoös, 12*(1), 5–15.

Sheinkopf, S. J., & Siegel, B. (1998). Home based behavioral treatment of young autistic children. *Journal of Autism and Developmental Disorders, 28,* 15–24.

Siegel, B. (1995). Assessing allegations of sexual molestation made through facilitated communication. *Journal of Autism and Developmental Disorders*, *25*(3), 319–326.

Silliman, E. R. (1995). "Issues raised by facilitated communication for theorizing and research on autism": Comment. *Journal of Speech and Hearing Research*, *38*(1), 204–206.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(8), 921–929.

Simpson, R. L., & Myles, B. (1995). Effectiveness of facilitated communication with children and youth with autism. *Journal of Special Education*, *28*(4), 424–439.

Smith, M., Haas, P. J., & Belcher, R. G. (1994). Facilitated communication: The effects of facilitator knowledge and level of assistance on output. *Journal of Autism and Developmental Disorders*, *24*(3), 357–367.

Smith, T., Eikeseth, S., Klevstrand, M., & Lovaas, O. I. (1997). Intensive behavioral treatment for preschoolers with severe mental retardation and pervasive developmental disorder. *American Journal of Mental Retardation, 102*, 238–249.

The Editors of the Lancet. (2010). Retraction—Ileal–lymphoid–nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet, 375*(9713), 445.

Tolbert, L., Haigler, T., Waits, M. M., & Dennis, T. (1993). Brief report: Lack of response in an autistic population to a low dose clinical trial of pyridoxine plus magnesium. *Journal of Autism and Developmental Disorders*, *23*(1), 193–199.

Tuzikow, J. E., & Holburn, S. (2011). Identifying fad therapies for autism spectrum disorders and promoting effective treatment. In J. L. Matson & P. Sturmey (Eds.), *International handbook of autism and pervasive developmental disorders* (pp. 307–319). New York: Springer.

Unumb, L. (2013, January). *Who pays for autism intervention, and who should?* Presentation given at the 5th annual meeting of the Council on Autism Services, Las Vegas, NV.

Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., et al. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet, 351*(9103), 637–641.

Wallis, C. (2006, May 10). "Helping" autistic people to speak. *Time U.S.* Available at *www.time.com/time/nation/article/0,8599,1192775,00.html.*

Wegner, D. M., Fuller, V. A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology*, *85*(1), 5–19.

Wermer, M. (2008). Dolphin therapy: The playful way to work toward the next step. *Exceptional Parent*, *38*(5), 70–72.

What is supported typing? (2010). Available at *http://soe.syr.edu/centers_institutes/institute_communication_inclusion/what_is_supported_typing/default.aspx.*

Wheeler, D. L., Jacobson, J. W., Paglieri, R. A., & Schwartz, A. A. (1993). An experimental assessment of facilitated communication. *Mental Retardation*, *31*(1), 49–59.

Wieder, S., & Greenspan, S. I. (2003). Climbing the symbolic ladder in the DIR model through floor time/interactive play. *Autism, 7*(4), 425–435.

Wieder, S., & Greenspan, S. I. (2005). Developmental pathways to mental health: The DIR model for comprehensive approaches to assessment and intervention. In K. M. Finello (Ed.), *The handbook of training and practice in infant and preschool mental health* (pp. 377–401). San Francisco: Jossey-Bass.

Williamson, C. (2008). Dolphin assisted therapy: Can swimming with dolphins be a suitable treatment? *Developmental Medicine and Child Neurology, 50*(6), 477.

Wong, H. H., & Smith, R. G. (2006). Patterns of complementary and alternative medical therapy use in children diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*, 901–909.

Zimmer, M. H. (2011). Complementary and alternative therapies for autism. In E. Hollander, J. T. Coyle, & A. Kolevzon (Eds.), *Textbook of autism spectrum disorders* (pp. 493–512). Arlington, VA: American Psychiatric Association.

CHAPTER FIFTEEN

Attachment Therapy

Jean Mercer

Attachment therapy (AT) is an unconventional mental health intervention that uses both physical and behavioral techniques to address certain emotional and behavior problems of childhood. These problems generally involve disobedience and unwillingness to express affection as desired by parents, although the treatment has been and still is used for autism. Most of the children treated have been adopted, but the timing of adoption may be anywhere from immediately after birth to school age or adolescence; older adopted children are likely to be foreign-born and to have had some period of institutional care.

Although it resembles other unconventional psychological treatments in some ways, AT is of particular interest because it has been embraced by popular culture more widely than most interventions. In addition, investigation of AT is important because the treatment has caused documented injuries and deaths. Focus on AT is also called for because most clients treated in this way are children or adolescents who are not in a position to refuse the intervention if their caregivers decide to use it.

AT has also been called Z-therapy (Zaslow & Menta, 1975), rage reduction therapy (Cline, 1992), holding therapy or holding time (Tinbergen & Tinbergen, 1983; Welch, 1989), prolonged parent–child embrace (Welch et al., 2006), or *Festhaltentherapie* (Prekop, 2008). Under all these names, the treatment involves physical restraint of a child by one or more practitioners. Following the death of a child in association with physical restraint, the treatment is said to have become less physically intrusive in nature, although holding the child in a cradling position remains a feature of the intervention. The present discussion will use the term attachment therapy (AT) because it is applicable whether or not a restraining hold is

used, and because it alludes to the aspect of personality AT practitioners believe themselves to be influencing. It should be noted that treatments with names like "attachment-focused therapy" are not necessarily the same as AT and may or may not use similar methods.

AT is commonly classed as a complementary-and-alternative (CAM) treatment because of its dearth of supporting evidence and because of its implausibility with respect to established positions on relevant aspects of child development. Because AT is neither "complementary" nor "medical" in the ordinary senses, the same reasons might more appropriately cause it to be called an alternative psychotherapy (AP), the term that will be used in the present paper. AT has also been categorized by some as a potentially harmful treatment (PHT) because of child deaths and injuries associated with it (Lilienfeld, 2007; Mercer & Pignotti, 2007).

THE NATURE OF AT

AT appears to exist in two major forms. The older one (AT1) involves both physical restraint and simultaneous provocation of the child's anger by tickling, poking, and shouting; this method generally starts with a calm child and works toward emotional distress, although a minor form is based on the claim that restraint of an emotionally aroused child has therapeutic value (Federici, 1998/2005; Ziegler, 2001). A second form (AT2) uses physical contact with a cradling hold that is said to offer nurture rather than restraint (Kelly & Popper, 2009). Both techniques emphasize "eye contact" (prolonged mutual gaze) as a therapeutic method (although AT2 does not force mutual gaze), and they share some other nonevidence-based tenets.

Diagnoses and Diagnostic Methods

Although AT1 has at times been used in attempts to treat autism (once proposed to be a disorder of attachment but not shown to be so [Sigman & Ungerer, 1984]), both AT1 and AT2 have generally been used in treatment of "attachment disorders" of one form or another, as the name of the treatment would suggest. Publications on AT often refer to the treatment as specific for reactive attachment disorder, a syndrome described in several versions of the *Diagnostic and Statistical Manual of Mental Disorders,* including DSM-5 (American Psychiatric Association, 2013). Reactive attachment disorder (RAD) was until recently diagnosed on the basis of several behavioral criteria describing developmentally inappropriate social behavior on the part of young children. It was considered to exist in an inhibited form, diagnosed when children are unusually shy and disengaged in social relationships, or a disinhibited form, in which they are unusually outgoing and friendly to strangers. These two forms, both called RAD, were the focus of AT1 and AT2 practice. DSM-5 separates the

two categories into reactive attachment disorder (the inhibited form) and disinhibited social engagement disorder. At the time of this writing, it is not clear whether AT1 and AT2 practitioners will claim treatment of the first, the second, or both disorders. Unusually for a DSM description, the criteria for RAD include an atypical history of social relations, including neglect, abuse, or frequent changes of caregiver. There is presently no well-accepted scale for diagnosis of RAD, although Minnis and others have proposed several approaches (e.g., Minnis, Rabe-Hesketh, & Wolkind, 2002).

Attachment Disorder

Although AT-related publications may reference the RAD diagnosis, and even quote the DSM criteria for the disorder, they may also refer to an extra-DSM diagnosis they call attachment disorder ("Symptoms, Causes, and Research," 2011). Even in cases where the RAD terminology is kept, the problem under consideration does not seem to be related to the DSM criteria. Attachment disorder (AD) is said to be characterized by an extensive list of disturbing symptoms such as fire-setting, "fascination with blood and gore," and "crazy lying" (lying that will obviously not be believed). These symptoms also include disobedience to parents and a failure to express affection "on the parent's terms" (see Buenning, 2012). For diagnosis of AD, the method recommended by AT proponents is the Randolph Attachment Disorder Questionnaire (RADQ; Randolph, 2000, 2001), an instrument completed by parents with the help of the therapist. Randolph has reported psychometric data for the RADQ, but has not made clear against what other measure of AD the RADQ has been validated; it would appear that the validation is against Randolph's own evaluation. Randolph has also stated that she is able to diagnose AD through observation of a child's ability to crawl backward on command. Sheperis et al. (2003) used the RADQ and a scale developed by Minnis et al. (2002) in efforts to create an office-based assessment.

With respect to diagnosis, Kelly and Popper (2009) more recently asserted on behalf of the Association for Treatment and Training of Attachment in Children (ATTACh) that "Reactive Attachment Disorder, as it is narrowly defined in the DSM-IV . . . has little to do with attachment patterns . . . and is largely unrelated to what is more commonly [*sic*] known as 'attachment disorder' " (p. 143). Kelly and Popper rejected the use of checklists for diagnosis and did not reference the RADQ (Randolph, 2000) or any associated work. Instead, they suggested that the disorders that bring children into ATTACh-approved treatment are the part of the complex or developmental trauma disorder proposed by van der Kolk (2005). This still-notional disorder contains within it a number of symptoms previously included on AD checklists and descriptions. For example, affected children are said to be hypersensitive to physical contact; to be analgesic; and to have difficulty localizing skin contact. No systematic

evidence has been presented for these symptoms as part of any relevant disorder, or for the rather startling claim that the children have difficulties with object constancy, an easily measurable perceptual ability. In any case, it is difficult to escape the conclusion that "attachment-focused therapy" diagnoses now have little to do with attachment as it is technically defined, as was the case for the older checklists.

Treatment Methods

Proponents of AT1 wrote and published detailed descriptions of their techniques and made training videos that were easily available to the public. In addition, 30 hours of videotape showing the treatment culminating in the death of Candace Newmaker were shown at the trial of her AT1 therapists (Mercer, Sarner, & Rosa, 2003). AT1 methods were also described by Stryker (2010) using observations she made in Evergreen, Colorado, in the late 1990s, at a peak of AT1 activity. With the advent of AT2, the method said to be less physically intrusive, a more corporate business model emerged. Access to examples or descriptions of AT2 techniques is now controlled to some extent by specific practitioners and by the Association for Treatment and Training of Attachment in Children (ATTACh), a parent–professional hybrid organization. As a result, although there are easily obtained detailed descriptions of AT1 methods, AT2 methods are less well understood by those outside the group that promulgates them.

AT1

The original, highly physically intrusive version of attachment therapy can involve two or more therapists who restrain a child, or (less commonly) an adolescent. The child's upper body lies in the lap of a seated primary therapist and is supported by the adult's nearer arm. The child's arm that is nearer to the therapist's body is placed behind the therapist, where the therapist can easily lean back on it or sit on the arm to immobilize it. A second therapist, seated nearby, holds the child's free hand. The primary therapist has a hand free to use in the physical actions that are part of the treatment.

In the course of the treatment, the primary therapist commands the child to kick with her feet, then to stop when told to do so. The child may also be ordered to get off the lap and to do push-ups or jumping jacks, then to return.When in the therapist's lap, the child is intermittently tickled, poked, and prodded in the ribs and under the arms by the therapist's free hand. At other times, the therapist holds the child's face firmly, insists on eye contact, and requires the child repeatedly to shout statements that reflect what the child is assumed to feel: "I hate my mother!"; "I want to kill you!" At intervals, the therapist talks quietly to the child, appears sympathetic, and comments on what a hard time the child has had. At other

times, the child is threatened both with abandonment by adoptive parents and by a prediction that he or she will kill someone and will go to prison.

This treatment continues for several hours at a time and may be repeated weekly or more often if the child is living at home. In an "intensive" treatment, the child may experience AT1 every day for two weeks and may be living in a therapeutic foster home (to be discussed later) while not in the treatment setting.

AT2

AT2 is said to have evolved from the original AT1 practices (Allred & Keck, n.d.). In a "White Paper" (*www.attach.org/WhitePaper.pdf*), the organization ATTACh rejected the use of coercion in treatment of childhood emotional disorders. In another ATTACh position paper (Kelly, 2003), treatment was described as involving a "corrective emotional experience" (Alexander & French, 1946) and particularly the need for experience of the empathic attunement of caregivers to the child's emotional state. This paper described attachment therapy as "an evolving, dynamic, heterogeneous set of interventions whose over-arching goal is the promotion of healthy attachment behaviors in families." In addition, with a discussion of the concepts of abreaction and catharsis, Kelly's paper recommended the release of negative emotions, but saw this as only a beginning of emotional expressivity, to be followed by pleasurable interactions. "Attachment therapy often seeks to provide concrete experiences of nurturance that allow regressive dependency needs to be met by a sensitive caregiver in the present," according to Kelly, who also argues that talk therapy cannot deal with the child's established defenses. Attachment therapy (AT2) thus includes

> touch and physical holding . . . to create a context that facilitates access to these defensively excluded components. First, physical closeness of the head in a parent's lap, coupled with encouraged eye contact, recreates a sense of dependency and vulnerability in the child. . . . more importantly, the protective lap . . . provides a physical experience of safety and comfort that can powerfully counter the potent sensorimotor memories of the early deprivation and maltreatment. (Kelly, 2003)

Kelly went on to note the presumed strong need of the child for control and suggested that this need makes it effective for parents to use "defiance-based paradoxical interventions." When the caregiver makes nurturing approaches to the child, the latter may make use of strategies of aggression in order to defend against the anxiety-laden situation. Physical holding is then considered to be a way of demonstrating to the child that these strategies do not work. However, Kelly pointed out that "in spite of the effectiveness of such defiance-based paradoxical interventions, their use requires special safety precautions" and has the potential for abuse.

In a 2009 ATTACh publication, Kelly and Popper referred to AT2 as "attachment-focused therapy" and repeated the earlier rejection of coercion, expressing concerns about re-traumatizing children. However, these authors also cited advocates of holding therapy (Keck & Kupecky, 1998; Levy & Orlans, 1998) as presenting possible therapeutic models. They noted the ATTACh position that the APSAC report (Chaffin et al., 2006) rejecting AT1 has some inaccuracies.

Adjuvant Treatments

Both AT1 and AT2 have been accompanied by adjuvant treatments and sometimes by milieu-therapy components. Of these, the rarest, but best-known, is *rebirthing*. This practice was publicized when it resulted in the death of Candace Newmaker in 2000 and has often been confused with AT1 itself. Rebirthing, a dramatization of a notional experience of one's own birth by movement through a constricted space, has been used primarily with adults in various AP settings. When used with children, it has usually been brief and harmless. In the Newmaker case, rebirthing appears to have been prolonged and intensified to the point of fatal asphyxiation, as a result of the AT1 therapists' belief that the child had to be forced to submit (see Shermer, 2004).

AT1 has generally been accompanied by *therapeutic foster home* or *therapeutic parenting* techniques. The foster parent and parent educator Nancy Thomas (2000) described these techniques in some detail. They include requiring the child to perform "strong sitting" or "power sitting"—sitting tailor-fashion without movement for periods of time, with the clock starting over each time there is movement. Thomas and others also recommended putting alarms on children's bedroom doors, removing furniture and lights from bedrooms, limiting food choices and amounts, refusing to answer children's questions about their parents, requiring that all privileges including toileting be asked for by the child, and demanding participation in tedious and pointless chores. In addition, Thomas suggested that at the convenience of the adult the child be rocked, hugged, and fed with milk-based sweets, all of which Thomas believes facilitate the development of an emotional attachment.

It is not clear whether AT2 also uses therapeutic parenting techniques. However, various AP methods, such as eye movement desensitization and reprocessing (EMDR), sensory integration (SI), and neurofeedback are recommended by ATTACh and by individual AT practitioners.

ATTACh statements about AT2 have been favorable to the use of dyadic developmental psychotherapy (DDP), a method that focuses on attunement of parents to a child's emotional state (Becker-Weidman & Hughes, 2008). Discussion later in this chapter will examine the relationship of DDP to both AT1 and AT2 and will consider whether DDP is more accurately classified as well or poorly evidence-supported.

HISTORICAL BACKGROUND OF AT

Tenets of AT can be traced to "wild psychoanalysts" (those who took many liberties with traditional psychoanalytic doctrine and practice) like Wilhelm Reich and to Freud himself. The specific principles and practices of AT1 and AT2 occupy more limited time frames, however. In this section, the *dramatis personae* and important events of the history of AT will be outlined. Other APs that have influenced AT will be referred to briefly. The early development of AT was simultaneous with Arthur Janov's primal therapy approach, with some of the rougher methods associated with transactional analysis (see Schiff, 1970), and with Synanon and other methods requiring personal submission.

AT as a definable set of practices began with the work of Robert M. Zaslow, a California psychologist and university faculty member, in the late 1960s and early 1970s. In a 1975 volume, Zaslow and his coauthor Marilyn Menta outlined a series of claims about the nature of attachment, its importance for autism and schizophrenia, and the interventions recommended. Citing the work of Reich, Zaslow described a method of physically restraining either children or adults (using large numbers of helpers to restrain the adults), tickling or prodding them, and forcing them to make eye contact. He believed that this method would create catharsis for rage that was blocking affectional behavior and that the release of rage would cure basic emotional problems.

Zaslow was not alone in this approach, as Martha Welch was beginning her "holding time" treatment, described by Tinbergen and Tinbergen (1983) and detailed by Welch in a 1989 book. Simultaneously, the Czech therapist Jirina Prekop adapted the Welch method, with the encouragement of Tinbergen and of the family therapist Bert Hellinger (2001). Welch continued to be active and visited the United Kingdom in the late 1980s on a tour in support of her book *Holding Time*.

Zaslow surrendered his psychology license following an incident in which an adult patient was injured. He demonstrated his methods around the country and eventually surfaced in Colorado, where he encountered the physician Foster Cline and instructed him on the "rage reduction" methods that would be systematized as AT1. Zaslow (1982) later reported that he had used his treatment on a girl at the Colorado School for the Blind, who regained her vision. Cline was the motivating force behind the development of clinical facilities performing AT in the town of Evergreen, Colorado. Producing some locally published books (1992) asserting that AT was the only appropriate treatment for children with reactive attachment disorders, defined by the checklist described earlier, Cline developed relationships with other AT proponents like the foster parent Nancy Thomas. At this time, Colorado did not require licensure or registration of psychotherapists; however, Cline surrendered his medical license following an injury to a child. He moved his base of operations to Idaho, a similarly unregulated

state, and established a parent-and-teacher education company under the name "Love & Logic."

The period of the 1990s was a highly active one for AT1, as described by Stryker (2010). The Association for Treatment and Training of Attachment in Children was incorporated not-for-profit in 1990 with 50 members, and began to offer conferences and training for both parents and mental health professionals, emphasizing AT1 as a desirable form of treatment and stating that conventional treatments exacerbated children's emotional problems. AT1 spread to distant parts of the United States and was promulgated by some state child welfare and social work organizations. In 1998, the Child Welfare League of America published a book promoting AT1 (Levy & Orlans, 1998), and in 2000 a second book edited by Levy was put out by Academic Press.

The death of Candace Newmaker in 2000, and the subsequent trial, conviction, and imprisonment of her therapists, caused serious reconsideration by ATTACh and AT1 advocates. ATTACh, which had supported AT1 as the only possible treatment for children with poor attachment histories, now posted position papers rejecting confrontational and coercive holding and proposing a "gentle nurturing" form of physical contact. However, ATTACh continues to claim that symptoms of attachment disorders include destructive behavior, fascination with blood and gore, and so on (see *www.attach.org*).

The group also altered its conference and training presentations, inviting major attachment disorder research figures like Charles Zeanah as keynote speakers. ATTACh began to offer continuing education (approved by the American Psychological Association and the National Association of Social Workers) that could culminate in a form of certification as a "registered attachment therapist" (of whom the ATTACh website presently lists 39) . Emphasizing AT2, ATTACh made an effort to enter the mainstream of mental health work, but did so without providing an evidence basis for the recommended treatment or contributing further to assessment efforts. Connections between ATTACh and major adoption groups have helped to maintain ATTACh's claimed position as the major source of attachment information for parents and provided a pulpit for AT2.

CONSIDERING THE EVIDENCE BASES FOR AT1 AND AT2

AP proponents frequently commit the error of critical thinking called "confident speculation"(Gula, 2002). Having developed a notional description of the mechanisms underlying mental illness, they accept their own speculation as established fact. From this error, they go on to establish interventions on the basis of the *pathophysiologic rationale*, a set of "inferences from (supposed) facts about the underlying pathological and physiological mechanisms of health and disease to conclusions that a treatment will or

will not have effects" (Howick, 2011, p. 16). To avoid these potentially mistaken inferences, a foundation of systematic evidence is essential, however difficult it may be to establish such a foundation for mental health interventions.

Informed Consent

Systematic collection of evidence for outcomes of AT methods is constrained by human subjects protection requirements under the Common Rule of the Department of Health and Human Services (HHS; "Federal Policy . . . ," n.d.). The 1983 subpart D of these regulations provides additional protection for children, whose parents must give permission for research participation, although direct permission is also required for older children and adolescents. HHS has recognized the potential for coercion of children (e.g., in the case of AT, through threats of abandonment if they do not engage in the treatment) or undue influence. The documentation of informed consent of both parents and children requires that they have been informed of alternate treatments and of the possibility of risks or discomforts. Without such documentation (which must be prospective), and without the approval of an institutional review board (IRB), research involving children is not appropriate, and journals often have policies of rejecting any submitted articles that have not met these human subjects protection criteria.

Most publications dealing with AT do not mention informed consent and neglect to report IRB submission and approval. The philosophy of AT1 did not support seeking the child's consent, but considered the exertion of adult authority to be a key factor in treatment. The informed consent document ("Informed Consent," 2011) posted by a practitioner of dyadic developmental psychotherapy, an AT2 form, does not mention either alternate treatments or potential risks or discomforts. The ATTACh advocate Victoria Kelly (Kelly, 2003) acknowledged the requirement of informed consent, but stated that

> severe trauma before the age of 4 results in limited cognitive recall and an inability to relate current symptoms to the early trauma. . . . Thus, the child, due to the very nature of the disorder, is unable to give informed consent. This is not an unusual occurrence with children given their limited cognitive abilities. Parents typically act on the child's behalf to give consent.

However, an incomplete document may not provide the information the parents need for truly informed consent. More recently, Kelly and Popper (2009) placed more emphasis on informed consent, but argued that children's emotional disorders may make them refuse treatment.)

Levels of Evidence

In spite of the problems of informed consent, several published reports address the efficacy of AT1 and AT2. It is not necessarily helpful to inquire whether these demonstrate the treatments to be evidence-based per se, but other questions may be useful. Mercer and Pignotti (2007) proposed a taxonomy of levels of evidence for treatments, including five possibilities: evidence-based interventions supported by randomized controlled trials and meeting other criteria such as replication; evidence-supported interventions with data from nonrandomized designs but meeting other requirements; evidence-informed interventions based on case studies or other weaker designs; belief-based interventions without published support or with support from composite cases only (often providing examples of the pathophysiologic rationale); and potentially harmful interventions with documented evidence of harm or an evident high probability of harm.

No empirical work on AT1 has been above the evidence-informed level. Lester (1997) used a simple pre–post research design, which is often an inappropriate approach to outcome studies of children because rapid developmental change in childhood is easily confounded with treatment, and modestly concluded that parents may have liked the AT1 treatment. Myeroff, Mertlich, and Gross (1999), in a paper based on Myeroff's 1997 dissertation, claimed that an AT1-treated group showed a decrease in symptoms compared with a no-treatment or community-treatment control group. However (as Myeroff acknowledged to some extent in her dissertation), there were a number of weaknesses in this nonrandomized study. An important one was the use of the RADQ as an outcome measure; empirical work on the RADQ has shown it to be uncorrelated with other diagnostic methods (Cappelletty, Brown, & Shumate, 2005). A second weakness was the use of a comparison group of families who had applied to bring their children to the treatment site in Evergreen, Colorado, but who had failed to do so for reasons that were unexplored, though perhaps relevant to the children's later conditions.

Welch et al. (2006) reported positive results from a study of prolonged parent–child embrace (PPCE), a physical restraint method that closely resembles AT1 in its insistence on child submission and its lengthy and intense application (see Tinbergen & Tinbergen, 1983; Welch, 1989). Like other AT1-related studies, Welch's work used the RADQ as one measure and noted that parents were more likely to complete that questionnaire than to complete the Achenbach Child Behavior Checklist (CBCL), which was also used. Welch et al. used normative data from the CBCL instead of a comparison group. All before-and-after comparisons were based on parent responses rather than other assessments, and the CBCL norms were also compared with parent responses. The authors noted that participating families had come from some distance for treatment and may have begun

with a bias in favor of a positive opinion; the possibility of regression to the mean was also mentioned.

There appear to be no published studies of AT2 per se, and as Kelly (2003) noted, there is no clear definition of this method. However, studies of the ATTACh-approved DDP have been reported, claiming that the therapy is "evidence-based" (Becker-Weidman, 2006a, 2006b; Becker-Weidman & Hughes, 2008). These have also been critiqued (Pignotti & Mercer, 2007; Mercer & Pignotti, 2007; Mercer, Pennington, Pignotti, & Rosa, 2009), with the conclusion that DDP cannot be described as evidence-based under any assessment protocol. Briefly, DDP studies have not used randomized controlled designs, they have employed the ill-researched RADQ as an outcome measure, and they have used inappropriate statistical techniques, for example, multiple *t*-tests rather than analysis of variance. In addition, material in one Becker-Weidman paper (2006b) suggests that rather than resembling DDP as presently described by its originator, Hughes (1998/2006), the method initially employed by Becker-Weidman shared important features with AT1.

EMDR, an adjuvant treatment associated with AT2, has no established evidence basis but has been marketed aggressively (Abeles & Koocher, 2011). SI therapy, a second associated adjuvant treatment, has received a review concluding that there is no evidence to support its principles or practices (Hyatt, Stephenson, & Carter, 2009). The therapeutic foster homes and therapeutic parenting associated with AT1 have never received evaluations separate from work examining the therapy.

Adverse Events

Consideration of the levels of evidence claimed for AT1 and AT2 is incomplete without examination of these interventions as potentially harmful treatments (Lilienfeld, 2007; Mercer & Pignotti, 2007). The occurrence of a number of adverse events in association with AT suggests that this category is the appropriate one for these forms of treatment. It is notable that neither minor physical injury nor harm to mental health is likely to be reported (although two AT1 practitioners, Foster Cline and Robert Zaslow, surrendered their professional licenses after injury to patients). Deaths are the most likely adverse events to be reported and clearly associated with a mental health treatment. Although it is possible that a small number of children have died, for example from an accidental fall, in the course of other mental health treatments (and even that therapists have died in their own offices), AT1 appears to be the only psychosocial intervention causally connected with patient deaths.

It should be noted that the physical techniques of AT1, including restraint, prodding, tickling, and so on, are not known to have caused deaths. Where deaths have occurred, they were not inevitable consequences of the intervention practices, but instead resulted when parents

or practitioners persisted inappropriately in a procedure because of their commitment to one of AT's principal tenets—that child submission to adult authority is central to improved mental health. Michael Shermer (2004) referred to this as "death by theory.")

The Candace Newmaker case (see Mercer, Sarner, & Rosa, 2003), mentioned earlier, was an obvious exemplar of the "death by theory" problem. The 10-year-old adopted girl was brought for treatment to Evergreen, Colorado, where AT1 was for some time a cottage industry. She was subjected to AT1 several times and stayed at a therapeutic foster home between sessions. Her death occurred in a "rebirthing" activity that was planned as a rest from therapy. The therapists, Connell Watkins and Julie Ponder, had recently attended a workshop on rebirthing given by an itinerant instructor, Douglas Gosney. They wrapped Candace in a flannel sheet that symbolized the womb and, together with a number of adult helpers, pressed on her wrapped body with pillows in imitation of the contractions of labor. Candace was instructed to "be born" by escaping from the sheet, but she was too tightly wrapped and could not do so. In spite of her vomiting and pleading for help, Watkins and Ponder kept the child in the wrap for 70 minutes, for the last 30 of which she was silent and unresponsive. Watkins and Ponder presumably made the assumption of other AT1 proponents (e.g., Reber, 1996), that vomiting or defecation and pleas for help were simply signs of resistance to change, which must be overcome by continuing assertion of authority.

Other deaths associated with AT1 have been caused by parents acting on the AT belief system or following instructions given them by therapists. Parents who went to trial for their acts sometimes attributed the child's death to his or her own acts, as in the case of David Polreis, a toddler who was argued to have beaten himself to death with a wooden spoon (Bowers, 2000). In this and other cases, AT1 proponents have assisted in mounting a "RAD defense." Defense attorneys argued that the child had injured himself intentionally as an act of aggression toward the parents, or that the parents had been forced to injure the child because of his behavior (Bowers, 2000). In other situations, it was argued that a history of attachment problems could be brought up with respect to an individual's later crimes (Niland, 2011).

Two cases involved apparent instructions from therapists that resulted in child deaths. A 4-year-old Utah adoptee, Krystal Tibbets, died when her adoptive father employed "compression therapy" and lay on her with all his weight, as a caseworker had insisted he do in spite of the child having stopped breathing in earlier sessions of the kind (personal communication, Alan Misbach). Another Utah child, Cassandra Killpack, died of hyponatremia when her adoptive mother forced her to drink a quantity of fluid as a paradoxical intervention when the child had taken a drink, disobeying the rule that she must request all food, drink, or other necessities (Warner, 2003). The mother was observed and helped by another child while forcing Cassandra to drink.

Other deaths of adoptees may also be linked to AT beliefs and practices, but multiple causes make it difficult to determine what happened. Viktor Matthey, a child adopted from Russia, died from the cumulative effects of exposure and beating, complicated by the impact of a diet of uncooked beans (Reilly & Hindash, 2002). Although features of the case resembled AT practices, this may also have been a matter of "common" or "garden variety" child abuse. Nathaniel Craver, a Russian adoptee, died following injuries that suggested restraint, beating, and starvation; an ATTACh member, Lark Eshleman, who had been working with the family, testified about attachment disorders in this case (Lanyon, 2011).

Records of minor injuries or of long-term psychological problems resulting from AT have never been investigated systematically, although they turn up sporadically (see Hafetz, 2000). One AT1 "survivor" initiated an Internet site for individuals who had been through AT1 as children, but all remained anonymous and no independent verification was possible. The present author interviewed a young woman who received AT1 (apparently in the Welch form) at her school as a young child. Following the break-up of a love affair, she developed a high level of anxiety and was preoccupied with the early experience, saying that she was most disturbed by the memory of the screams of other children in the group treatment room. She was treated at a university anxiety clinic and is doing well. A British case in which AT1 is reported to have been used by a major child welfare organization is presently under investigation.

It is difficult to tell whether abuse cases reported by journalists or tried in courts involve AT1 beliefs, but it is possible that when these result in severe weight loss there may be some AT1 connection. This speculation is strengthened when therapists who have been AT1 proponents are shown to have been involved with the children's families, as appeared in *State v. Salvetti* (2010), a case in which an adopted child was confined to a bedroom and given a limited diet for several months. Following his escape, the parents served a prison sentence, but the therapist was not charged.

THEORY, RESEARCH, AND PLAUSIBILITY

Mental health interventions whose outcomes have never been evaluated adequately may nevertheless have promise. Whether such promise exists can be assessed to some extent by considering the plausibility of the intervention and its congruence with accepted tenets of mental health treatment. To do this, we need to examine the theories and principles on which an intervention relies, and to determine whether the claimed theoretical background is foundational to a treatment. Professionals are likely to ignore implausible treatments without supporting evidence, although there is a remote possibility that their implausibility means that a process has been wholly misunderstood in the past. The absence of supporting evidence for

AT makes an examination of its plausibility an important step, particularly because AT proponents have claimed that lack of evidence should not be taken as evidence against the effectiveness of their method (Kelly & Popper, 2009).

Congruence with Established Theories

Association with Attachment Theory

Both AT1 and AT2 proponents state that their interventions are based on attachment theory as developed by John Bowlby (1982). The use of the term "attachment" and the assumption that early social interactions establish the trajectory of personality development are congruent with Bowlby's work and with a great deal of later evidence (Sroufe, 2005). One of Bowlby's early papers (1944) suggested a connection between a poor attachment history and later delinquent behavior. However, none of Bowlby's work assumed that child obedience, gratitude, and displayed affection for caregivers are necessarily measures of attachment, or that changes in attachment will bring about these characteristics in older children. Bowlby's description of events in the development of attachment emphasized the beginning of attachment behavior in the second half of the first year of life, in contradiction to the belief of AT proponents that all adopted children, even if adopted shortly after birth, suffer from disorders of attachment. Although Bowlby discussed the possibility that juvenile thieves might be influenced by difficulties with attachment, he never suggested that extreme violence (e.g., serial killing) was associated with attachment problems, as AT advocates have been known to do (Fallahi, n.d.). Bowlby's statement that adult attachment figures are ordinarily stronger and wiser than the children who become attached to them was never elaborated by him into the AT-related belief that submission to authority is the cause of attachment (Cline, 1992). In short, attachment therapy is implausible in terms of what is known about the nature of attachment and its outcomes.

Ethological Concepts

AT1 theories were supported several decades ago by Nikolaas Tinbergen, the ethologist Nobel Laureate, and some AT concepts are plausible within the ethological theoretical context. Tinbergen and Tinbergen (1983) were concerned with the nature and treatment of autism, a disorder on which Martha Welch was at that time focusing her physical restraint methods. Tinbergen argued that innate releasing mechanisms were at work in creating a learned emotional connection between adults and their young, as appears to be the case for the phenomenon of imprinting in some birds. As Bowlby had done at one time, Tinbergen suggested a parallel between imprinting and human emotional attachment. As would be the case for a duck

deprived of early experience with moving objects to follow, human infants were thought to fail in attachment if they did not experience a releaser that would trigger the social preference. Tinbergen made some efforts to identify supernormal releasers, like a mask with large eyes, that might attract the attention of autistic children and initiate innate attachment behaviors (Tinbergen, 1974). When Welch proposed that face-to-face restraint was responsible for releasing attachment behavior, Tinbergen enthusiastically accepted this idea, which was quite plausible within the ethological framework. Work showing that autistic children do not lack attachment (Sigman & Ungerer, 1984) and that attachment is a robust phenomenon that can develop in a range of ways lessens the plausibility of these ideas.

Reichian Theory

AT1 is closely related to the theories of Wilhelm Reich (1945), a psychoanalyst who proposed that personality is associated with muscular tensions. Reich's therapeutic methods included the prodding of the trunk and armpits characteristic of AT1 and were referenced by Robert M. Zaslow (Zaslow & Menta, 1975), who proposed the use of physical restraint and linked autism and other childhood mental disorders to attachment problems. Reich's views and methods were sufficiently known in the mid-20th century to be described in a popular abnormal psychology textbook (White, 1948), but they were never adopted by mainstream mental health approaches; like AT, they have remained incongruent with established systems of treatment.

Psychoanalytic Assumptions

Two other aspects of AT were derived from beliefs still current in psychoanalytic circles but implausible with respect to present-day psychological thought. The first of these is catharsis, the discharge of negative emotion through reexperience of a disturbing situation, a concept incongruent with the work of Littrell (2009), but essential to AT in both its forms (Cline, 1992, 1994; Kelly, 2003). The second and more complex concept at work in AT is the assumption of regression and recapitulation of development. The regression assumption states that under certain circumstances, an individual's present-day mental organization can be caused to revert to an earlier life stage; the circumstances may be a reenactment of earlier experiences like physical contact with a parent or like being fed. When regression is caused, an alteration in the developmental trajectory is said to result, so that the individual quickly reorganizes at a higher level, leaving behind any problems that might have resulted from poor early experiences and achieving a more mature and positive mental state. That regression in this sense can occur at all is highly questionable, and the claim that recapitulation can correct emotional problems is unsupported (Mercer, 2011a, 2011b). These two foundational aspects of AT1 and AT2 are implausible.

Ericksonian Influence

Kelly's (2003) paper discussing the principles of AT2 refers to an important theme, the paradoxical and other approaches of Milton Erickson and of Jay Haley (1990). Erickson's work was of particular interest to Foster Cline, one of the initial promulgators of AT1. In a book influential among AT proponents (Cline, 1992), one of Erickson's papers is reprinted in sections, with the source cited only in the first section, so that a careless or naïve reader could easily confuse the Erickson work with the Cline text. The Erickson (1962) paper is an account of a case in which a divorced mother sought help with her noncompliant preteen son; the mother wanted to date, and the son was resisting this change. Erickson advised the mother to sit on her son for hours at a time and to prepare for this process by having food, drink, and reading material for herself within reach. Following the session, the mother was to provide cold oatmeal for the son, while preparing appetizing food for herself and another child. This regime was to be diminished as the son became more compliant, and Erickson concluded that it was effective, noting with apparent approval that after treatment the boy trembled when his mother spoke to him. Erickson's work in this case seems to have provided a template for AT advocates. However, the congruence of these treatment methods with actions listed as neglectful or abusive for research purposes (Sedlak, Mettenburg, Schultz, & Cook, 2005) suggests that the treatment is implausible with respect to accepted methods.

Haley's paradoxical intervention approach was directed toward helping a client gain voluntary control over impulsive behavior like public masturbation. AT1 practitioners may have interpreted their insistence on a child's shouting that he hated his mother as a paradoxical attempt to gain control over negative emotions, but their discussion of this matter suggests that they were attempting catharsis. The use of paradoxical interventions in conjunction with AT appears to have been distorted into "making the punishment fit the crime." For example, in the case of Cassandra Killpack, mentioned earlier, forced fluid consumption was the response to her disobedient drinking—an implausible application of the paradoxical intervention technique.

The "Attachment Cycle"

An essential tenet of AT theory involves the existence of an "attachment cycle," a series of experiences in the first year of life that are said to foster the attachment of a child to a parent—a process AT authors emphasized in spite of their stated belief that attachment has already occurred before a child is born. The theoretical importance of the attachment cycle for AT proponents appears to lie in the possibility it offers of regression and recapitulation of personality development. Without some concept of postnatal development of attachment, it would be much more difficult to imagine a

reenactment of attachment events involving some form of prenatal experience only.

Vera Fahlberg, whose ideas influenced AT thinking in the 1990s, described an "attachment cycle" or "arousal-relaxation cycle" (Fahlberg, 1990, pp. 33–34), which she considered to be essential to early attachment. In this cycle, the infant progresses from a quiescent, satisfied state to the experience of a need and a sense and communication of displeasure. Caregivers respond to this and return the child to the quiescent state. Many repetitions of the cycle "help[s] the child to develop trust, security and to become attached to his primary caregiver" (Fahlberg, 1991/2012, p. 34). Fahlberg attributed some part of this belief to Spitz (1965). The attachment cycle has repeatedly been referenced by AT proponents, including Welch (Welch et al., 2006), who uses a diagram similar to Fahlberg's to describe the events she considers to occur during her PPCE therapy. According to Welch (2006, p. 10), repeated embraces that cycle through distress and end in calm acceptance are necessary for effective treatment:

> Both mother and child quickly learn or relearn to associate the comfort of synchronous attunement in each other's embrace with the accompanying stress-free state of homeostasis or calm arousal. After many PPCE nurturing cycles [periods of restraint—my addition], the child's response to stress is positively conditioned and the child is able to maintain calm arousal for longer and longer periods of time. With repeated achievement of a synchronous and modulated state, the child also develops empathy. He or she learns to give as well as receive comfort.

This treatment, according to Welch et al. (2006), "appears to replicate or reinstate the effects of normal early nurturing."

The "attachment cycle" or "nurturing cycle" is plausible in terms of early Freudian explanations of attachment that identified attachment as due to "cupboard love," or gratification through feeding experiences. But these cycles are much less plausible in terms of present-day thinking about social and personality development in the first year of life. From Bowlby's time forward, personality development has been described in terms of social interactions. Greenspan (Greenspan & Wieder, 2006) described in detail the types of social experiences related to steps in attachment. Practical and ethical barriers to experimental investigation of these theories have limited work in this area, but observations suggest that feeding experiences, though essential to life and development, are a minor factor in attachment. For instance, the boarder baby phenomenon (Maza, 1999) argues against the idea that regular care is the issue in personality development. Boarder babies are infants who are kept in a hospital because, although healthy, they have no homes to go to. They are fed and cared for regularly and well, and many people interact with them cordially, though briefly. Their physical development is good, but by 8 or 9 months they tend to show a

"shallow" social response characterized by failure to carry through the long chains of social interactions normally expected by this time.

Trauma Concepts

The theory behind AT2 has been described by Kelly (2003) in an ATTACh position paper cited earlier. Two major theoretical tenets underlie AT2. The first is an emphasis on trauma as a precursor of attachment disorders (as defined by this group). Following Perry (2000) and Briere (2002), AT2 assumes that early trauma is experienced and remembered at an implicit level. Initial treatment also needs to be at this implicit level to "decondition the automatic fear reactions associated with emotional intimacy" (Perry, 2000). Physical holding is recommended as a way to get to these automatic reactions, an idea that is plausible within the context of current theories of trauma.

A second basic tenet of AT2 is that events known to facilitate infant social and personality development also facilitate improvement in mental health (see, for example, Becker-Weidman & Hughes, 2008). Thus, the interactive mismatch-and-repair (Tronick, 1989) characteristic of developing communication between healthy infants and their caregivers is assumed to play a role in treatment of older children. This tenet is shown at a cruder level in the recommendations of AT practitioners that older children be bottle-fed with warm, sweetened milk. These beliefs in recapitulation of early experiences are implausible in terms of current understanding of development, in which traumas or deficits are understood to alter developmental trajectories rather than to halt development at some point from which it can be induced to proceed.

Congruence with Relevant Research

Adoption as Problematic

An essential topic to examine with respect to the congruence of AT with mainstream work is adoption. Although the initial use of AT1 addressed autism, the focus quickly moved to adoption as a cause of reactive attachment disorder. AT proponents have stressed the importance of their treatments for all adopted children, including those adopted soon after birth) and have claimed that all such children experience disorders of attachment that will culminate in serial killing or prostitution. However, these claims are implausible when considered in the context of existing research.

Several studies have confirmed that the long-term differences in emotional development of adopted and nonadopted children are small. For example, Sharma, McGue, and Benson (1998) concluded that there was little evidence that adoption was associated with later violent behavior. Children adopted before age 6 months from orphanages have been shown

to catch up effectively in all areas of development (Croft et al., 2007), contradicting the AT belief that all adopted children have problematic futures.

Most saliently, two longitudinal studies of children living in or adopted from orphanage conditions have shown the implausibility of the AT position. One of these studies, by the St. Petersburg–USA Orphanage Research Team (2008), considered development in the early years in institutions that were adequate in terms of medical and physical care but lacking in social and emotional interactions with adult caregivers. The outcomes of caregiver training for improved social interactions with children were examined. Although the children remained in the group care that AT proponents regard as pathogenic, children associating with trained caregivers became more emotionally expressive, sought more proximity and contact with caregivers, and were less likely to be categorized as disorganized/disoriented in their attachment behavior than children whose caregivers received different or no training. All of these findings render the AT view of the grief and rage associated with separation from the birth mother rather implausible. (This is not, however, to deny some of the behavior problems found among children adopted from deficient orphanages; the information simply suggests that the AT belief system is mistaken about the causes of those problems.)

Another longitudinal study, the English and Romanian Adoptees (ERA) study (Rutter et al., 2010), investigated deprivation-specific psychological patterns in children adopted from severely depriving institutions in the 1980s and early 1990s. This group of children was not only separated from their birth mothers but had been neglected and abused, a state of affairs that AT proponents would predict should lead to severe attachment disorders and a high probability of violent behavior. So far, the study has followed more than 100 children adopted from Romanian orphanages up to the age of 15 years. Kumsta et al. (2010), in their chapter in the Rutter volume, reported their analysis of deprivation-specific patterns (DSPs) that occurred in early life for children who had lived in socially depriving institutions, but were rare in other groups. To be identified as a DSP, the pattern would have to be distinctively different from common mental health problems of childhood, as well as meeting other criteria. Kumsta et al. put forward a possible DSP that emerged in about 15% of the ERA children, but in less than 1% of a pooled comparison group. This pattern involved four factors, with the first being quasi-autism, an impaired ability to pick up social cues and understand social contexts. The quasi-autistic children also had "unusual, intense, circumscribed interest patterns and unusual preoccupations showing a similar obsessive-like quality" (p. 54). The second factor in this possible DSP was the existence of disinhibited attachment, a category of reactive attachment disorder as described by DSM-IV-TR and a separate disorder as described by DSM-5. These children are indiscriminately friendly, appear unaware of social rules about physical boundaries, and may seek physical closeness to an unusual extent. The third and fourth factors in the possible DSP were cognitive impairment and a tendency to

inattention and overactivity. Notably, however, when the ERA children reached their teens, many became less conspicuously quasi-autistic, and in some cases the inappropriate friendliness became a strength, as the adolescents were seen as socially outgoing (Kreppner et al., 2010).

AT positions are incongruent with the ERA studies. AT1 proponents have argued that all adopted children, and many nonadopted individuals, will develop attachment disorders, even though they show no symptoms at a given time. In addition, it has been predicted that such children, if not treated with AT, will become violent "psychopaths" and "serial killers." The ERA studies show that the great majority of children adopted from socially deprived, neglectful, even abusive institutions develop well within normal limits. For those whose development is atypical, there is nevertheless no resemblance to the characteristics that AT practitioners claimed for adopted children. In addition, even the atypically developing ERA children often did better with age, whereas the AT view is that adolescence is likely to be a time of worsening disturbance and violence. Finally, the operative AT belief is that problems of adopted children arise from disturbances of attachment. Gunnar (2010), commenting on the ERA work, cited concerns with the assumption that so-called disinhibited attachment is actually a matter of attachment.

Comparison to Pentecostal Practices

AT principles thus appear to be incongruent with established views of personality change and understanding of child development, except in that they share the current concern with the impact of trauma. AT1, at least, may be more congruent with an observable belief system outside psychology than with mainstream psychological thought. The system in question is a set of beliefs termed "deliverance" by Pentecostal Christians. The essence of the deliverance system is the belief that all physical and mental disorders are caused by demonic possession. Although some AT proponents, such as Nancy Thomas, accept this belief, it would be absurd to suggest that AT is based on this idea. However, assumptions and methods of AT1 and of deliverance practitioners overlap to a remarkable extent. Both are concerned that separation from the birth mother and subsequent adoption (even if occurring at a very early age) are important causes of mental illness. Both focus on child obedience and affection toward caregivers as indices of mental health. Both use firm commands and physical restraint as techniques of treatment, and both are concerned with sexual acting-out and violent behavior as consequences of untreated problems. Both consider "eye contact" to be indicative of mental health status. Finally, both consider vomiting or coughing up mucus to be an expected response to treatment (Bialecki, 2011; Cuneo, 2001). It would be speculative to claim that AT emerged from the deliverance system, but the homologies are impressive.

PROFESSIONAL, LEGISLATIVE, AND JUDICIAL RESPONSES TO AT

AT beliefs and methods developed "under the radar" for some years, but after about 2000 a number of individuals and groups stated objections to AT itself and to some of its associated outcomes.

Professional Concerns

Only one child mental health professional, Beverly James, stated clear opposition to and alarm over AT before 2000, the year of Candace Newmaker's death. James (1994) described "holding" methods in detail but cited no sources, making it difficult for readers to explore the matter further. Candace's death drew particular interest because of the intense discussion then going on about the use of restraint and seclusion in mental health treatment (Gross, Mitchell, & Hayes, 2003). Several published papers in the next few years commented on the problems associated with AT (Mercer, 2001, 2002a, 2002b; Kennedy, Mercer, Mohr, & Huffine, 2002), followed by a book discussing the Newmaker case (Mercer et al.; Sarner, & Rosa, 2003). A special issue of the journal *Attachment and Human Development* (2003) included articles rejecting AT by O'Connor, Steele, Zeanah, and other leaders of attachment disorder research, as well as one by Daniel Hughes (once an AT1 proponent) and one by Minnis and Keck, the last named long associated with AT1 techniques (see Keck & Kupecky, 1998). Barth, Crea, John, Thoburn, and Quinton (2005) argued against the overemphasis on attachment issues for adoptees and in favor of evidence-based interventions.

Candace Newmaker's death and the trials of her therapists were the impetus for statements from a number of professional societies, repudiating AT1, but in some cases confusing the treatment with rebirthing. A 2005 congressional resolution ("Urging prohibition of rebirthing techniques," 2005) rejected the use of rebirthing, calling it an attachment therapy technique and encouraging states to adopt legislation prohibiting it, while referencing a statement of the American Psychological Association that rejected rebirthing as an appropriate form of treatment. (This statement followed an earlier positive/neutral description of AT methods by DeAngelis [1997] in the American Psychological Association *Monitor*.) The American Psychiatric Association rejected the use of coercion or restraint as therapeutic methods, as did the National Association of Social Workers, the American Academy of Child and Adolescent Psychiatry, and the British Association for Adoption and Fostering. Most notably, the American Professional Society on Abuse of Children convened a task force (which included an ATTACh official) and released a report discussing in detail the reasons for rejection of AT1 methods (Chaffin et al, 2006).

Legislative Actions

Legislation prohibiting AT1 methods has been difficult to achieve. Not only have groups of mental health professionals objected to legislative management of their professional activities, but it is difficult to describe a protocol in such a way that it cannot easily be changed in some detail by those who wish to continue its use. Colorado enacted "Candace's Law" in 2001, prohibiting the use of rebirthing techniques that involve restraint. The fact that the child's death occurred in Colorado, and the presence of anti-AT advocates in that state, were factors in this successful, though incomplete, legislation. In 2002 and again in 2003, the Utah state legislature considered a bill to ban AT1 practices. The unsuccessful legislation was supported by virtually every mental health organization and child advocacy group in the state and received broad national support, but was opposed by "pro family" activists who argued that the government should not dictate the treatment that parents can choose for their children. Utah State Representative Mike Thompson, who sponsored the anti-AT1 legislation, lost his bid for reelection.

Judicial Findings

Criminal charges against parents who used AT1 methods have gone to court in several cases, but the charges involved injuries to the children that were classified under laws about abuse and neglect; the AT1 methods are not in themselves illegal. For example, the charges against Sylvia Jovanna Vasquez in a 2006 case concerned her confining several of her adopted children to cages and restricting their diets to bread, water, and peanut butter. Vasquez attributed her actions to the advice of a book by Nancy Thomas, an AT1 advocate, which had been recommended by an adoption caseworker. Vasquez was imprisoned for a time during the pretrial phase because of her attempts to communicate with the children, but eventually she was sentenced to little more than time served. In a similar case, Michael and Sharen Gravelle, an Ohio couple who kept their 11 adopted children in cages, were convicted on felony counts of child endangerment and misdemeanor counts of child abuse. The two oldest children later filed a suit for damages against the counseling agency that had characterized the Gravelles as suitable adoptive parents. The family attachment therapist, Elaine Thompson, was indicted on charges of failure to report child abuse.

In a fatal case in Pennsylvania, Michael and Nanette Craver were convicted of involuntary manslaughter, endangering the welfare of a child, and criminal conspiracy following the death of their child adopted from Russia (Lanyon, 2011). Their therapist, Lark Eshleman, a member of ATTACh, authored a 2003 book that lists among "typical features of RAD" a number of items like lying, hoarding, and destructive behavior, characteristic

of the AT viewpoint; the volume also recommends "planned regressions." Evidence about the actual methods used by the Cravers was difficult to establish, but a forensic pathologist rebutted the claim that the child had essentially killed himself, and noted that the body showed signs of severe starvation.

AT AND VERNACULAR PSYCHOLOGY

The story of AT is to some extent a story of popular culture. Although the emergence of AT in the 1970s occurred in academic/professional settings, the boundaries between those settings and popular thought were porous during that period, as psychologists and others explored "ancient wisdom," traditional practices, and drugs as ports of entry to mental processes. The 1964 Old Saybrook conference on humanistic psychology had conferred approval on this approach in the names of Abraham Maslow, Rollo May, Carl Rogers, and other influential psychologists. Although professional psychology soon turned back to encouragement of a more science-oriented system, and by the 1990s began an emphasis on evidence-based treatment, popular culture has continued to define psychology quite broadly and to focus on "easy-to-understand" AP concepts rather than on the complexities of outcome research. Popular culture has embraced AT specifically in several ways, creating a vernacular psychology that is reflected in the mass media and that increasingly influences some levels of academic and professional activity.

Professional, legal, and judicial reactions against AT appear impressive but mean little in terms of impact on vernacular psychological beliefs. Unexamined concepts of a folkloric nature (Bennett, 1987) are supported by interactions among informal groups and may only be strengthened by the opposition of the authorities. An example of the influence of groups holding AT-related vernacular beliefs is the circulation of a "letter to teachers" (Murphy, 2002) advising them of the importance of conforming to AT assumptions about children's lying and the primacy of parental authority.

When AT support is easily found in the mass media, members of the public are likely to find related testimonial statements far more impressive than the views of the "experts." Googling "attachment disorders adopted children treatment" yields 2,480,000 results, of which many focus on AP methods, including AT1, AT2, EMDR, neurofeedback, sensory integration therapy, and so on. (The relevant Wikipedia articles do not have this bias, due largely to the efforts of the present writer, another Wikipedia editor, and a sympathetic administrator.) These pages offer testimonials or poorly designed research as evidence that their treatment "works."

In addition to web pages that promote AT methods, but that presumably are intentionally searched for by interested parties, the mass media present general-interest material that supports AT principles and methods.

Motion pictures that offer such support began with the Elvis Presley 1969 vehicle *Change of Habit*, in which Presley took the role of a physician who treated an autistic child to the tune of screams off-camera, and as a result enabled her not only to speak but to say that she loved her mother. A 1989 HBO "documentary" entitled "Child of Rage" is still in circulation and available on Youtube (*www.youtube.com/watch?v=g2-Re_Fl_L4*). This production features the supposed life story of Beth Thomas, the adopted daughter of the AT1 advocate Nancy Thomas. "Child of Rage" focuses on the violent, disturbing behavior of the young child, and her admirable development following AT1 methods. A 2012 independent documentary, "My Name Is Faith" (*www.hotdocs.ca/film/title/my_name_is_faith*), purports to tell the story of an adopted child with "detachment disorder," with the AT1 foster parent Nancy Thomas representing a therapist. Unsurprisingly, there are no systematic depictions of the potential harm done by AT or of the success of evidence-based methods of treatment.

The possible effects of media presentations on parents who are seeking help with what they believe are attachment disorders may be multiplied by the influence of related state-sanctioned training of mental health and welfare workers, and of foster and adoptive parents. Arizona, Georgia, New Mexico, and Pennsylvania are among the states that have provided such training. Support for AT has also come from some religious organizations; Focus on the Family has posted discussions of reactive attachment disorder that refer to the attachment cycle and the work of Cline (Grebenik, n.d.). National professional organizations have also been supportive of continuing education about AT principles and practices. Approved providers of CE credits for the American Psychological Association, National Association of Social Workers, and marriage and family therapists' groups have all given credits for AT-related presentations, thus encouraging practitioners to accept views that are congruent with popularly supported approaches.

AT BELIEFS, EDUCATION, AND PUBLICATION

Is the AT system gradually moving from a vernacular level to a more formal academic and professional level? Acceptance of AT beliefs in educational settings and in scholarly publishing would suggest that this is happening.

Acceptance in Educational Settings

At the time of Candace Newmaker's death, few educational settings provided or supported training or research on AT principles or practices. The Union Institute of Ohio, an intermittently-accredited organization, had granted doctorates to several AT proponents, including Robin Myeroff, whose 1997 dissertation on the efficacy of AT1 has often been referenced by AT advocates. A 2001 master's thesis from Virginia Polytechnic

Institute emphasized AT1 views. A 2009 Tennessee doctoral dissertation used the RADQ as an outcome measure. A 2010 Columbia College (Chicago) dance and movement therapy master's paper focused on RAD and cited Reber's 1996 article, a paper frequently used to support AT1 beliefs. Capella University awarded doctorates for dissertations using AT concepts in 2005 and 2007. The University of Wisconsin at Stout awarded a master's degree for a thesis with a strong AT orientation. Some of these degrees were in psychology, some in social work, and some in guidance, as well as one in dance therapy. Incidentally, the majority of the relevant undergraduate term papers offered for sale on line appear to be AT1-related.

Changes in Publications

Examining the publication of AT-oriented books also suggests a movement toward acceptance of AT into the mainstream. The original AT1 publication (Zaslow & Menta, 1975) was brought out by the San Jose State University Press, but in the form of a duplicated typescript rather than a typeset volume. From the 1980s to the present, AT material shifted from publication by "printer-ready" presses to the involvement of more rigorous and mainstream publishers. The formerly self-published ATTACh professional practice guide (Kelly & Popper, 2009), for example, was published by W.W. Norton in 2012 (Becker-Weidman, Pessolano Ehrman, & LeBow, 2012).

Third-Party Payment

An additional measure of the intrusion of AT beliefs into popular culture may be provided by the willingness of insurance companies to cover AT procedures. Naturally, these enterprises are less than enthusiastic about coverage of unusual services and have adopted guidelines favoring evidence-based treatment. However, some states reimburse through Medicaid for AT and provide AT as a postadoption service.

Support by Tradition

Consideration of the influence of popular culture on acceptance of AT also raises questions about traditions of child discipline as they merge into abusive treatment. Historically, common practices resembling AT, such as confining a child to a dark closet, withholding food, demanding immobility and tedious work, and putting soap or other unpleasant substances into a child's mouth, have only fairly recently been categorized as abusive. Groups that claim adherence to traditional values still advocate some of these methods, as well as the use of intense physical punishment for minor disobedience of infants and children (Ezzo & Bucknam, 1998; Pearl & Pearl, 1994). The AT belief that children can control tantrums, defecation,

and vomiting, and that these acts are sometimes carried out as intentional defiance of authority is a traditional assumption and is reflected in some child abuse cases. For example, descriptions of the 2012 abuse death of Khalil Wimes in Philadelphia (Newall, 2012) refer to punishments meted out for vomiting and soiling, as well as withholding of food and punishment for "stealing" food. The AT perspective may seem familiar and therefore acceptable to psychologically unsophisticated parents who recall discipline methods used by their own parents or by neighbors and relatives. Popular beliefs can thus influence both parents and practitioners to ignore information about the nature of child abuse and to pursue AT methods that should be categorized as abusive. In turn, the popularization of AT methods may lessen the likelihood that child welfare workers will recognize actions as abusive.

The Adoption Subculture

Beliefs of an adoption subculture have also been supportive of AT principles and practices. A document of the Evan B. Donaldson Adoption Institute (Evan B. Donaldson, 2010) describes some nonevidence-based, AT2-related interventions as if they are evidence-based, and references ATTACh (as well as more suitable sources of information). Bethany Christian Services, an organization that provides adoption and postadoption services throughout much of the United States, lists a library that includes material by Cline, Randolph, Thomas, and other AT1 proponents (Bethany Christian Services, 2012) and notes that the library was funded by the group AdoptUSKids. Dozens of adoption blogs refer approvingly to AT beliefs and methods, potentially strengthening impressions about the prevalence of attachment-related problems among adopted children, and the need for unusual treatments—even stating, as AT proponents have often done, that conventional interventions exacerbate child mental health problems.

International Events

AT has generally been considered an intervention native to the United States, but AT methods have burgeoned internationally. Treatment with AT has been reported to have been used by a major British child-welfare organization (Chaika, 2012). The Czech holding therapist Jirina Prekopova, after years of practicing an AT method in Germany, has returned to the Czech Republic and had some success in popularizing her techniques as a treatment for autism and for children's oppositional behavior (as shown at *www.ceskatelevize.cz/porady/10265737370-mali-tyrani/210542151030001*). Prekopova has also opened clinics in Spain and Latin America. The AT parenting practitioner Nancy Thomas was invited to speak in Russia by a neo-Pentecostal group (see *www.globalstrategy.ru/News/jan_2012/25_01_12.html*).

CONCLUSION

The principles and practices of AT are not plausible with respect to established theory and research, nor are they based on systematic evidence meeting stringent criteria. Nevertheless, AT beliefs have been embraced by popular culture and are communicated daily through the Internet and other media. Such beliefs, which are easy for the layperson to find and understand, may be taking the place of more complex scientific information about the nature of emotional attachment and of childhood mental health disorders. Metaphorically, AT may be seen as an invasive species choking the growth of relevant evidence-based views among the general population, or perhaps as a form of debased currency—bad "money" that drives out the good. Like chiropractic treatment or homeopathy in competition with evidence-based medicine, AT appears to the layperson to be more understandable, more accessible, cheaper, and less stigmatizing than evidence-based child mental health interventions.

What is to be done? The first step, implemented in the present volume and its predecessor, is to educate clinicians about the existence of APs and their appeal to the public. Many psychologists have never heard of AT in any of its forms and are unaware that parents they work with may already be committed to this unconventional system. Without that awareness, clinicians cannot know what those parents want and expect from treatment, or what factors may move them to avoid evidence-based practices.

A second step is to focus on public education about early emotional development, which surveys of parents (Zero to Three, 2000) have shown to be the least understood aspect of development. Instructors in child development courses have a serious responsibility in this area, as do those who teach "family" courses in high schools—the latter using curricula that have benefited little or not at all from advances in psychological understanding.

In spite of these and other possibilities, it is difficult to avoid pessimistic conclusions about the continuing influence of AT. Folkloric systems are by definition unexamined by their believers, and it is hard to see why AT should be any exception.

GLOSSARY

Attachment theory: A framework for understanding infants' development of strong preferences for familiar caregivers and subsequent developmental changes in social relationships.

Attachment therapy: One of several forms of child psychotherapy focused on creating an intense emotional attachment to a caregiver in a child past infancy; these treatments usually involve some form of physical contact and may include coercive restraint. Attachment therapists attribute many mood and behavioral problems to disorders of attachment.

Ethology: A framework for the study of animal behavior that focuses on unlearned responses and on readiness for rapid social learning in early life; a theoretical source for a form of holding therapy.

Holding therapy: One of at least two types of treatment involving coercive restraint of a child, often by a parent; used in attempts to treat autism, oppositional behavior, and problems attributed to disorders of attachment.

Potentially harmful treatment: A psychotherapy that has caused, or could possibly cause, unwanted events ranging from failure to find effective treatment to exacerbation of symptoms to physical injury or death.

Reactive attachment disorder: A set of behaviors characterized by young children's atypical conduct toward familiar caregivers, following a history of early neglect or abuse.

Rebirthing: A treatment involving reenactment of some aspects of the posited experiences of an infant during birth, especially of movement through a narrow passage; claimed to correct for the effects of difficult birth or to create intense emotional attachment to a caregiver.

Vernacular psychology: A set of beliefs about human development and behavior that are common outside professional and academic psychology; often without empirical foundation, these beliefs may determine the acceptance of some alternative treatments.

"**Wild psychoanalysis":** A term used by Sigmund Freud to describe schools of treatment that emphasized reenactment of early experiences and sometimes used physical contact as part of treatment.

REFERENCES

Abeles, N., & Koocher, G. P. (2011). Ethics in psychotherapy. In J. C. Norcross, G. R. Vandenbos, & D. K. Freedheim (Eds.), *History of psychotherapy: Continuity and change* (2nd ed., pp. 723–740). Washington, DC: American Psychological Association.

Alexander, F., & French, T. (1946). *Psychoanalytic therapy: Principles and application*. New York: Ronald Press.

Allred, D., & Keck, G. (n.d.). The evolution of attachment therapy. Available at *www.abcofohio.net/treatment.htm*.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. Arlington, VA: Author.

Attachment and Bonding Center of Ohio. (n.d.). Application process. Available at *www.abcofohio.net/application.htm*.

Barth, R. P., Crea, T. M., John, K., Thoburn, J., & Quinton, D. (2005). Beyond attachment theory and therapy: Toward sensitive and evidence-based interventions with foster and adoptive families in distress. *Child and Family Social Work, 10,* 257–262.

Becker-Weidman, A. (2006a). Treatment for children with trauma-attachment disorders: Dyadic developmental psychotherapy. *Child and Adolescent Social Work Journal, 23,* 147–171.

Becker-Weidman, A. (2006b). Dyadic developmental psychotherapy: A multi-year follow-up. In S. M. Sturt (Ed.), *New developments in child abuse research* (pp. 43–60). New York: Nova Science.

Becker-Weidman, A., & Hughes, D. (2008). Dyadic developmental psychotherapy: An evidence-based treatment for children with complex trauma and disorders of attachment. *Child and Family Social Work, 13,* 329–337.

Becker-Weidman, A., Pessolano Ehrman, L. A., & LeBow, D. (2012). *The attachment therapy companion: Key practices for treating children and families.* New York: Norton.

Bennett, G. (1987). The rhetoric of tradition. *Talking Folklore, 1*(3), 32–46.

Bethany Christian Services. (2012). Library. Available at *https://secure.bethany.org.*

Bialecki, J. (2011). No caller ID for the soul: Demonization, charisms, and the unstable subject of Protestant language ideology. *Anthropological Quarterly, 84,* 679–704.

Bowers, K. (2000, July 27). Suffer the children. Available at *www.westword.com/content/printVersion/215282.*

Bowlby, J. (1944). Forty-four juvenile thieves: Their character and home-life. *International Journal of Psychoanalysis, 25,* 19–52.

Bowlby, J. (1982). *Attachment and loss. Vol. 1: Attachment.* New York: Basic Books. (Original work published 1969)

Briere, J. (2002). Treating adult survivors of severe childhood abuse and neglect: Further development of an integrative model. In J. E. B. Myers, L. Berliner, J. Briere, C. T. Hendrix, T. Reid, & C. Jenny (Eds.), *The APSAC handbook on child maltreatment* (pp. 175–202). Newbury Park, CA: Sage.

Buenning, W. D. (2012). RAD checklists. Available at *www.reactiveattachmentdisordertreatment.com/childattachchecklist.pdf.*

Cappelletty, G. G., Brown, M. M., & Shumate, S. E. (2005). Correlates of the Randolph Attachment Disorder Questionnaire (RADQ) in a sample of children in foster placement. *Child and Adolescent Social Work Journal, 22,* 71–84.

Chaffin, M., Hanson, R., Saunders, B. E., Nichols, T., Barnett, D., Zeanah, C., et al (2006). Report of the APSAC Task Force on Attachment Therapy, Reactive Attachment Disorder, and Attachment Problems. *Child Maltreatment, 11,* 76–89.

Chaika, A. (2012). *Invisible England.* London: CreateSpace.

Cline, F. W. (1979). *What shall we do with this kid?: Understanding and treating the disturbed child.* Evergreen, CO: EC Publications.

Cline, F. W. (1992). *Hope for high risk and rage filled children.* Evergreen, CO: EC Publications.

"CRB training module." (2003). Available at *www.nmcrb.org*

Croft, C., Beckett, C., Rutter, M., Castle, J., Colvert, E, Groothues, C., et al. (2007). Early adolescent outcomes of institutionally-deprived and non-deprived adoptees: II. Language as a protective factor and a vulnerable outcome. *Journal of Child Psychology and Psychiatry, 48,* 31–44.

Cuneo, M. W. (2001). *American exorcism: Expelling demons in the land of plenty.* New York: Doubleday.

DeAngelis, T. (1997). When children don't bond with parents. *American Psychological Association Monitor.* Available at *www.toddlertime.com/dx/rad/attachment-parenting.htm.*

Dziegielewski, S., & Forbes, H. (2002). Reactive attachment disorder. In S. Dziegielewski (Ed.), *DSM-IV-TR in action* (pp. 143–168). Hoboken, NJ: Wiley.

Erickson, M. H. (1962). The identification of a secure reality. *Family Process, 1*(2), 294–303.

Eshleman, L. (2003). *Becoming a family: Promoting healthy attachments with your adopted child*. Lanham, MD: Taylor.

Evan B. Donaldson Adoption Institute. (2010). *Keeping the promise: The critical need for post-adoption services to enable children and families to succeed*. Available at *www.adoptioninstitute.org/publications/2010_10_20_KeepingThePromise.pdf*.

Ezzo, G., & Bucknam, R. (1998). *On becoming babywise*. Sisters, OR: Multnomah.

Fahlberg, V. I. (1990). *Residential treatment: A tapestry of many therapies*. Indianapolis, IN: Perspectives.

Fallahi, C. R. (n.d.). Reactive attachment disorder. Available at *www.psychology.ccsu.edu/fallahic/RAD%20outlines.pdf*.

Fattah, G. (2002, September 17). Parents charged in Springville girl's death. *Salt Lake City Deseret News*. Available at *www.deseretnews.com/article/937544/Parents-charged-in-Springville-girls-death.html?pg=all*.

"Federal Policy for the Protection of Human Subjects (Common Rule)." (n.d.). Available at *www.hhs.gov/ohrp/humansubjects/commonrule/index.html*.

Federici, R. S. (1998/2005). *Help for the hopeless child*. Alexandria, VA: Author.

Grebenik, D. (n.d.). A new definition of attachment/regulation. Available at *www.focusonthefamily.com*.

Greenspan, S. I., & Wieder, S. (2006). *Infant and early childhood mental health*. Arlington, VA: American Psychiatric Publishing.

Gross, G., Mitchell, B., & Hayes, A. (2003). Restraint and seclusion: Overview of federal laws and policies. Available at *www.hogg.utexas.edu/uploads/documents/federal_laws%20.pdf*.

Gula, R. J. (2002). *Nonsense: A handbook of logical fallacies*. Mt. Jackson, VA: Axios Press.

Gunnar, M. R. (2010). A commentary on deprivation-specific psychological patterns: Effects of institutional deprivation. In M. Rutter et al (Eds.), Deprivation-specific psychological patterns: Effects of institutional deprivation. *Monographs of the Society for Research in Child Development, 73*(1, Serial No. 295), 232–247.

Hafetz, D. (2000, March 11). Parents charged with denying food, medical care to 7-year-old. *Austin American-Statesman*, p. B1.

Haley, J. (1990). *Strategies of psychotherapy*. New York: Norton.

Hellinger, B. (2001). *Love's own truths: Bonding and balancing in close relationships*. Phoenix, AZ: Zeig, Tucker, & Theisen.

Howick, J. H. (2011). *The philosophy of evidence-based medicine*. New York: Wiley-Blackwell.

Hughes, D. (1998/2006). *Building the bonds of attachment*. Northvale, NJ: Aronson.

Hyatt, K. J., Stephenson, J., & Carter, M. (2009). A review of three controversial educational practices: Perceptual motor programs, sensory integration, and tinted lenses. *Education and Treatment of Children, 32*(2), 313–342.

"Informed consent." (2011). Available at *www.center4familydevelopment.com. informedconsent.htm*.

James, B. (1994). *Handbook for treatment of attachment-trauma problems in children*. New York: Simon & Schuster.

Keck, G., & Kupecky, R. (1998). *Adopting the hurt child: Hope for families with special-needs kids*. Colorado Springs, CO: NavPress.

Kelly, V. J. (2003). Theoretical rationale for the treatment of disorders of attachment. Available at *www.attach.org/theorational.htm*.

Kelly, V. J., & Popper, S. (Eds.). (2009). *Attachment-focused therapy: A professional practice guide*. Lake Villa, IL: ATTACh.

Kennedy, S., Mercer, J., Mohr, W., & Huffine, C. (2002). Snake oil, ethics, and the First Amendment: What's a profession to do? *American Journal of Orthopsychiatry, 72*(1), 5–15.

Kreppner, J., Kumsta, R., Rutter, M., Beckett, C., Castle, J., Stevens, S., et al (2010). Developmental course of deprivation-specific psychological patterns: Early manifestations, persistence to age 15, and clinical features. In M. Rutter et al. (Eds.), Deprivation-specific psychological patterns: Effects of institutional deprivation. *Monographs of the Society for Research in Child Development, 73*(1, Serial No. 295), 79–101.

Kumsta, R., Kreppner, J., Rutter, M., Beckett, C., Castle, J., Stevens, S., et al. (2010). Deprivation-specific psychological patterns. In M. Rutter et al. (Eds.), Deprivation-specific psychological patterns: Effects of institutional deprivation. *Monographs of the Society for Research in Child Development*, *73*(1, Serial No. 295), 48–78.

Lanyon, A. (2011). Trial continues in adopted Russian boy's death. Available at *www.wset.com/story/15457730/trial-continues-in-adopted-russian-boys-death*.

Lester, V. S. (1997). Behavior change as reported by caregivers of children receiving holding therapy. Available at *www.attach.org/lester.htm*.

Levy, T. M. (2000). (Ed.). *Handbook of attachment interventions*. San Diego, CA: Academic Press.

Levy, T. M., & Orlans, M. (1998). *Attachment, trauma, and healing*. Washington, DC: Child Welfare League of America.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2,* 53–70.

Littrell, J. (2009). Expression of emotion: When it causes trauma and when it helps. *Journal of Evidence-Based Social Work, 6,* 300–320.

Lyman, L. (2001). Reactive attachment disorder: A training outline. Available at *www.pacwcb.pitt.edu*.

Maza, P. (1999). Boarder babies and placement in foster care. *Clinics in Perinatology, 26,* 201–211.

Mercer, J. (2001). "Attachment therapy" using physical restraint: An object lesson on the identification of unvalidated treatments. *Journal of Child and Adolescent Psychiatric Nursing, 14*(3), 105–114.

Mercer, J. (2002a). Child psychotherapy involving physical restraint: Techniques used in four approaches. *Child and Adolescent Social Work Journal, 19*(4), 303–314.

Mercer, J. (2002b). Attachment therapy: A treatment without empirical support. *Scientific Review of Mental Health Practice, 1*(2), 9–16.

Mercer, J. (2011a). The concept of psychological regression: Metaphors, mapping, Queen Square, and Tavistock Square. *History of Psychology, 14,* 174–196.

Mercer, J. (2011b). Some aspects of CAM mental health interventions: Regression, recapitulation, and "secret sympathies." *Scientific Review of Mental Health Practice, 8,* 36–55.

Mercer, J., Pennington, R., Pignotti, M., & Rosa, L. (2009). Dyadic developmental psychotherapy is not "evidence-based": Comments in response to Becker-Weidman and Hughes (2009). *Child and Family Social Work, 15,* 1–5.

Mercer, J., & Pignotti,M. (2007). Shortcuts cause errors in systematic research syntheses: Rethinking evaluation of mental health interventions. *Scientific Review of Mental Health Practice, 5*(2), 59–77.

Mercer, J., Sarner, L., & Rosa, L. (2003). *Attachment therapy on trial*. Westport, CT: Praeger.

Minnis, H., Rabe-Hesketh, S., & Wolkind, S. (2002). Development of a brief, clinically relevant, scale for measuring attachment disorders. *International Journal of Methods in Psychiatric Research, 11*(2), 90–98.

Murphy, J. (2002). Reactive attachment disorder: A summary for teachers. Available at *www.attachmentnewengland.com/reactive_attachmet_disorder.pdf.*

Myeroff, R., Mertlich, G., & Gross, G. (1999). Comparative effectiveness of holding therapy with aggressive children. *Child Psychiatry and Human Development, 29,* 303–313.

Newall, M. (2012, April 24). How city agencies failed to save Khalil. *Philadelphia Inquirer,* pp. 1, 10–11.

Niland, J. (2011). Reactive attachment disorder. Available at *http://voiceforthedefenseonline.com/story/reactive-attachment-disorder.*

Pearl, M., & Pearl, D. (1994). *To train up a child*. Pleasantville, TN: No Greater Joy Ministries.

Perry, B. D. (2000). The neurodevelopmental impact of violence in childhood. In D. Schetky & Benedek (Eds.), *Textbook of child and adolescent forensic psychology* (pp. 221–238). Washington, DC: American Psychiatric Press.

Pignotti, M., & Mercer, J. (2007). Holding therapy and dyadic developmental psychotherapy are not supported, acceptable social work interventions. *Research on Social Work Practice, 17*(4), 513–519.

Prekop, J. (2008). *Ich halte dich fest, damit du frei wirst*. Munchen: Kosel-Verlag.

Randolph, E. (2000). *Manual for the Randolph Attachment Disorder Questionnaire*. Evergreen, CO: Attachment Disorder Press.

Randolph, E. (2001). *Broken hearts, wounded minds*. Evergreen, CO: RFR Productions.

Reber, K. (1996). Children at risk for attachment disorder: Assessment, diagnosis, and treatment. *Progress: Family Systems Research and Therapy, 5,* 83–98.

Reich, W. (1945). *Character analysis*. Rangely, ME: Orgone Institute Press.

Reilly, M., & Hindash, S. (2002). The short life of Viktor Matthey. Available at *www.dartcenter.org/content/short-life-viktor-matthey.*

Rutter, M., Sonuga-Barke, E. J., Beckett, C., Castle, J., Kreppner, J., Kumsta, R., et al (2010). Deprivation-specific psychological patterns: Effects of institutional deprivation. *Monographs of the Society for Research in Child Development*, Serial No. 295, *75*(1), 1–252.

Schiff, J. L. (1970). *All my children*. New York: M. Evans.

Sedlack, A., Mettenburg, J., Schultz, D., & Cook, D. (2005). *NIS definitions*

review. Rockville, MD: Westat. Prepared for the United States Department of Health and Human Services.

Sharma, A., McGue, M., & Benson, P. (1998). The psychological adjustment of United States adoptive adolescents and their non-adopted siblings. *Child Development, 69,* 791–802.

Sheperis, C. J., Doggett, R. A., Hoda, N. E., Blanchard, T., Renfro-Michael, R.I., Holdiness, S. H., et al. R. (2003). The development of an assessment protocol for reactive attachment disorder. *Journal of Mental Health Counseling, 25*(4), 291–310.

Shermer, M. (2004, June). Death by theory. *Scientific American*. Available at *www.scientificamerican.com/article.cfm?id=death-by-theory*.

Sigman, M., & Ungerer, J. A. (1984). Attachment behaviors in autistic children. *Journal of Autism and Developmental Disorders, 14*(3), 231–244.

Spitz, R. A. (1965). *The first year of life: A psychoanalytic study of normal and deviant development of object relations*. New York: International Universities Press.

Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. *Attachment and Human Development, 7,* 349–367.

State v. Salvetti (2010). Available at *www.caselaw.findlaw.com/nc-court-of-appeals/1503115.htm*.

St. Petersburg-–USA Orphanage Research Team (2008). The effects of early social-emotional and relationship experience on the development of young orphanage children. *Monographs of the Society for Research in Child Development, 73*(3, Serial No. 291).

Stryker, R. (2010). *The road to Evergreen*. Ithaca, NY: Cornell University Press.

"Symptoms, causes, and research." (2011). Available at *www.attachmentdisorder.net/Symptoms_Causes_Research.htm*.

Thalman, J. (2003, January31). House rejects holding therapy. Salt Lake City Deseret News. Available at *www.deseretnews.com/article/962238/House-rejects-holding-therapy.html?pg=all*.

Thomas, N. (2000). Parenting children with attachment disorders. In T. M. Levy (Ed.), *Handbook of attachment interventions* (pp. 67–111). San Diego, CA: Academic Press.

Thomson, L. (2004, June 2). Center is sued over therapies. Salt Lake City Deseret News. Available at *www.deseretnews.com/article/595072140/Center-is-sued-over-therapies.html?pg=all*.

Tinbergen, N. (1974). Ethology and stress diseases. *Science, 185,* 20–27.

Tinbergen, N., & Tinbergen, E. (1983). *"Autistic" children: New hope for a cure*. Boston: Allen & Unwin.

Tronick, E. Z. (1989). Interactive mismatch and repair: Challenges to the coping infant. *Zero to Three, 6,* 1–6.

"Urging prohibition of rebirthing techniques." (2005, October 7–26). *Congressional Record, 151* (Pt. 17).

Van der Kolk, B. (2005). Developmental trauma disorder. *Psychiatric Annals*, *35*, 374–378.

Warner, L. (2003, October 9). Killpacks face trial in death of daughter. *Salt Lake City Deseret News*. Available at *www.deseretnews.com/article/515037467/Killpacks-face-trial-in-death-of-daughter.html?pg=all*.

Welch, M. G. (1989). *Holding time.* New York: Fireside.

Welch, M. G., Northrup, R. S., Welch-Horan, T. B., Ludwig, R. J., Austin, C. L., & Jacobson, J. S. (2006). Outcomes of prolonged parent–child embrace therapy among 102 children with behavioral disorders. *Complementary Therapies in Clinical Practice, 12*(1), 3–12. Available at *www.marthawelch.com.*

White, R. W. (1948). *The abnormal personality.* New York: Ronald Press.

Wimmer, J. S., Vonk, M. E., & Bordnick, P. (2009). A preliminary investigation of the effectiveness of attachment therapy for adopted children with reactive attachment disorder. *Child and Adolescent Social Work Journal, 26*(4), 351–360.

Zaslow, R. W. (1982). Der Medusa-Komplex. Die Psychopathologie der menschlichen Aggression in Rahmen der Attachment-Theorie, widergespielt in Medusa-Mythos, dem Autismus und der Schizophrenie. *Zeitschrift fuer Klinische Psychologie und Psychotherapie, 30*(2), 162–180.

Zaslow, R. W., & Menta, M. (1975). *The psychology of the Z-process: Attachment and activity.* San Jose, CA: San Jose State University Press.

Zero to Three. (2000). What grownups understand about child development: A national benchmark survey. Available at *www.zerotothree.org.*

Ziegler, D. (2001). To hold, or not to hold . . . Is that the right question? *Residential Treatment for Children and Adolescents, 18*(4), 33–45.

CHAPTER SIXTEEN

Antisocial Behavior of Children and Adolescents

Harmful Treatments, Effective Interventions, and Novel Strategies

Anthony Petrosino, Pamela MacDougall, Meghan E. Hollis-Peel, Trevor A. Fronius, and Sarah Guckenburg

On February 7, 2013, two fifth-grade boys were arrested in Colville, Washington, for plotting to kill a female classmate whom they found "annoying" (Eng, 2013). According to news reports, the boys brought a gun, bullets, and a large knife to school. Their plot involved one boy stabbing the girl while the other boy held off staff and other students with the firearm. The newspaper accounts also indicate a plan to murder six other students, one at a time. Both boys were arrested for attempted murder.

In the aftermath of the plot, parents, educators, police, courts, and mental health professionals were forced to grapple with what to do with these boys. It would be wonderful if a plethora of effective evidence-based treatments were available to choose from. The unfortunate truth is that many treatments provided to children who engage in antisocial behavior are not effective, and some may be harmful. Moreover, interventions that exert no or minimal effects may result in opportunity costs—that is, waste of time, money, or missing possibilities to try something else that may be effective. In fact, the history of research in this area underscores the crucial point that some of the anticipated "cures" developed for antisocial behavior of children actually cause more harm than good (e.g., McCord, 2003).

In this chapter, we examine treatments for antisocial children. After defining what antisocial behavior is and some of the complexities involved in identifying it, we provide an overview of the scope of the problem and highlight theories of how youth become antisocial. We then categorize treatments of antisocial behavior of youth into three categories: (1) ineffective or harmful treatments, (2) effective interventions, and (3) novel or untested programs. The chapter concludes with recommendations for policy and research.

Our chapter considers interventions only for antisocial behavior among children, and not adults, although some treatments may be effective for both. By children, we mean children in the United States approximately ages 5–17, years encompassing the time period between youth and adulthood (age 18), going from kindergarten to 12th grade. Additionally, our chapter focuses on responses to antisocial behavior in general, particularly acts that come to the attention of the authorities, rather than how to treat youth with specific diagnoses such as conduct disorder or antisocial personality. Although there is overlap in both groups, sometimes substantially so, certain youth who commit delinquent acts would not be formally diagnosed with a mental health disorder. Moreover, some youth have received such formal diagnoses but have not committed acts that would rise to a criminal or legal level, or otherwise come to the attention of authorities. Finally, we focus only on intervention programs, or programs and policies that respond to antisocial acts already committed. Readers should note that a number of prevention programs designed for implementation at preschool, primary/secondary school, or larger community level are designed to reduce childhood antisocial behavior.

DEFINITIONS AND FORMAL DIAGNOSES

Antisocial behavior can be defined in a variety of ways. For example, the United Kingdom's Crime and Disorder Act of 1998 defined antisocial behavior as behaving "in a manner that caused or was likely to cause harassment, alarm, or distress to one or more persons not of the same household." Kazdin (1987) wrote, "Antisocial behaviors include a broad range of activities, such as aggressive acts, theft, vandalism, fire setting, lying, truancy, and running away. Many terms—including acting out, externalizing behaviors, conduct disorder, or conduct problems, and delinquency—denote antisocial behaviors" (p. 187). Antisocial behavior encompasses a wide variation of conduct that violates either legal codes or social norms (Rutter, Giller, & Hagell, 1998). Eddy and Reid (2001) defined "antisocial behavior" as a constellation of related behaviors, including disobedience, aggression, temper tantrums, lying, stealing, and violence.

Although many people may commit an antisocial act (though perhaps not a crime) at some point in their lives, most researchers would agree

that a youth is not antisocial until he or she engages in antisocial behavior consistently and over an extended period of time (Moffitt, 1993). Eddy and Reid (2001) noted that some antisocial behaviors are typical at certain stages of child development (e.g., the toddler throwing his toys at a parent), but it is the combination of antisocial acts, especially during adolescent years, that signals later adjustment problems, including criminal behavior in older teens and adults.

The psychiatric diagnosis (DSM-5; American Psychiatric Association, 2013) of conduct disorder includes such behaviors as physically harming or threatening people and animals, destruction of property, lying or theft, and repeated violations of rules such as running away from home or being truant from school. Conduct disorder is further diagnosed based on the age of onset. Individuals with symptoms of conduct disorder prior to age 10 are diagnosed with childhood-onset type, which is the type most likely to lead to a diagnosis of antisocial personality disorder in adulthood. Individuals with antisocial behavior with no symptoms of conduct disorder prior to age 10 are diagnosed with adolescent-onset-type conduct disorder and are less likely to exhibit symptoms of conduct disorder throughout adulthood (American Psychiatric Association, 2013).

Children identified with conduct disorder can also be diagnosed with antisocial personality disorder as they age. According to DSM-5 (American Psychiatric Association, 2013), an individual is eligible for a diagnosis of antisocial personality disorder when he or she is at least 18 years of age, has a history of conduct disorder, and displays a pattern of chronic disregard for the law and the safety of others, including at least three of the following behaviors: failure to conform to societal norms of appropriate behavior, dishonesty, poor impulse control, guiltlessness, hostility or aggressiveness, irresponsibility, and wanton lack of concern for others' safety; these behaviors cannot better be accounted for by schizophrenia or the manic episodes of bipolar disorder.

As mentioned earlier, the legal system is concerned with the actual behavior of the juvenile. For example, a youth may become engaged with the juvenile justice system for committing an antisocial act but may not have been formally diagnosed with a personality or mental health condition such as conduct disorder. Moreover, many of the behaviors that would concern psychological professionals, such as lying, deceitfulness, lack of remorse, and irritability, would not necessarily rise to the level of concern for the legal system.

THE PROBLEM

Although cases as severe as the Colville, Washington, incident are rare, antisocial behavior in childhood is not. Persistent and serious forms of antisocial behavior are estimated to be perpetrated by 5–10% of children in

developed Western countries (Scott, Doolan, Beckett, Harry, Cartwright, et al., 2012). Assuming that this percentage covers youth in the United States, an estimated 3,350,000–7,770,000 youth ages 17 and younger would have been considered antisocial in 2013 (Child Stats, 2013).

After conducting a meta-analysis of longitudinal studies in the United States, Lipsey and Derzon (1998) found antisocial behavior displayed by children ages 6–12 was a major predictor of serious or violent offending in the older teen and young adult years. Antisocial behavior as a child serves as the strongest predictor of criminal behavior during adulthood (Eddy & Reid, 2001). For example, Eddy and Reid (2001) indicated that a sizable percentage (40–75%) of youth who are arrested for a crime or meet the psychiatric criteria for conduct disorder are arrested in adulthood. Youth who become involved in delinquency at young ages are at great risk for committing crime in adulthood and experiencing other problems, such as struggling at school and getting involved in drugs or risky sexual behavior (Eddy & Reid, 2001).

The societal consequences of antisocial youth who are not successfully treated can be enormous. For example, one estimate puts the costs to the public of a single antisocial youth who continues on a law-breaking path to be approximately $1.7–2.3 million due to crime, public assistance, loss of wages, and harm to others over his or her lifetime (Cohen, 1998). Thus, antisocial behavior hurts not only the youth involved, but others in the family, school, or larger community and costs society millions of dollars.

Given the considerable personal and societal costs of antisocial acts, it is not surprising that government agencies and professionals would be concerned with reducing such behaviors. For example, in the United States, legislatures at the state level have passed laws to enhance penalties for youth who commit certain offenses, and agencies and nonprofits have developed programs to address antisocial behavior. Antisocial behavior is the most common issue confronted by mental health professionals. For example, Kazdin (1995) reported that antisocial behavior accounts for 33 to 50% of youth who are referred for treatment. Moreover, guidance counselors and school administrators often interact with antisocial youth and are required to take action when children persistently engage in harmful or disruptive behavior in schools. Finally, police, probation officers, and staff at juvenile reformatories all encounter youth who have participated in antisocial behavior.

PREDICTING ANTISOCIAL BEHAVIOR AND SELECTED THEORIES

Knowledge of the exact pathway for antisocial behavior in children remains elusive. Nevertheless, research provides insights into risk and protective factors and provides fodder for potential explanatory theories. Kazdin et al. (1997) defined risk factors as variables that predict the increased

probability of later offending behaviors. Still, the term "risk factor" has not been used consistently in the literature (Farrington & Welsh, 2007, p. 17). Protective factors, in contrast, are often defined as being on the "opposite end of the scale to a risk factor" (Farrington & Welsh, 2007, p. 23). In other words, protective factors predict a decreased likelihood of future offending or buffer the impact of risk factors. Rutter (1985) defined a protective factor as any moderator variable that interacts with one or more risk factors to minimize its effects on the outcome of interest.

Risk and protective research studies often examine what factors in children's background or experience are correlated with subsequent criminality. The strongest studies are longitudinal, follow a cohort of youth (delinquent and nondelinquent), and assess characteristics or traits of interest on one occasion and antisocial behaviors at a later time (Petrosino, Derzon, & Lavenberg, 2009). Because many children engage in antisocial acts, particularly as smaller children, and because it is difficult to identify the "onset" of antisocial behavior, especially when the youth is engaging in such actions on a continual basis, researchers often use delinquency or crime as an outcome measure (Farrington & Welsh, 2007).

Researchers have examined a wide range of risk and protective factors for antisocial behavior. Individual-level risk factors comprise those that arise from the interaction between the individual and his or her environment (Farrington & Welsh, 2007). Key individual-level risk factors include low intelligence and achievement, lack of empathy, impulsiveness, poor social-cognitive skills (Farrington & Welsh, 2007), and a mental disorder (e.g., Washburn et al., 2007). For example, Loeber et al. (2003) found that youth with conduct disorder are 17 times more likely to develop antisocial personality disorder, compared with individuals with no conduct disorder diagnosis. Researchers have also linked conduct disorder to parents' criminality (genetic influence), unemployment, substance abuse, poor physical health, and mental disorders (Odgers et al., 2007).

Farrington and Welsh (2007) also identified family-level risk factors. They found that the best predictors of child antisocial behavior (in order from most to least strong) were poor parental supervision, parental rejection, large family size, low parental involvement, parental conflict, and antisocial or criminal parents (Farrington & Welsh, 2007). Additional risk factors classified as socioeconomic included economic deprivation, whereas risks factors associated with peers included membership in a delinquent or deviant peer group, delinquent or antisocial friends, peer rejection, and gang affiliation. Farrington and Welsh (2007) identified school risk factors, including distrust between teachers and students, lack of commitment of the child to school and conventional academic goals, and unclear or inconsistently enforced school rules. Finally, they identified community risk factors, including areas of residence characterized by physical deterioration, neighborhood disorganization, and high residential mobility.

Although many theories have been posited to explain why individuals

commit crime, we describe a few here that are most relevant to children and antisocial behavior. Developmental life-course theory attempts to explain crime across all life stages (Sampson & Laub, 2005). According to this theory, becoming involved in crime is the product of a developmental process that starts before birth and continues throughout the life course. Individual and environmental risk and protective factors interact to determine the onset, length, and end of criminal careers. A focus of researchers from the development life-course theoretical perspective is on why individuals continue to engage in crime and why they desist or drop out of crime altogether.

Moffit (1993) also put forward a developmental theory of criminal behavior. She delineated two types of offenders: (1) the adolescent-limited offender, who exhibits antisocial behavior only during adolescence; and (2) life-course-persistent offenders, who begin antisocial behavior in childhood and continue through adulthood. Moffit (1993) argued that theories of onset and continuance of crime differ for these disparate groups. Most of her theory is focused on life-course persistent offenders, who present the larger crime problem. She proposed that biological and environmental risk factors, including brain injuries, parental rearing, and genetic vulnerabilities interact to explain this developmental path.

Low self-control or impulsivity theory (Gottfredson & Hirschi, 1990) emerged in the 1990s and continues to be popular. Control theorists contend that criminal motivation is widespread, and the presence or absence of controls leads to crime. In earlier formulations, external controls, such as relationships or social bonds, were regarded as protective factors. In the latest iteration of control theory, Gottfredson and Hirshi (1990) argued that youth become criminal because they do not develop "self-control" or the ability to restrain their impulsivity. Accordingly, key risk and protective factors involve parenting and other societal and institutional supports for teaching children to resist impulsive behavior. Left unrestrained, impulsive youth will commit "acts of force or fraud" to satisfy immediate gratifications.

Theories attempt to integrate risk and protective factors into a coherent explanation of behavior. Still, as noted above, children who engage in antisocial behaviors impose a burden on society. Although government and societal institutions strive to curb antisocial behaviors with laws, programs and services for youth, and family interventions, some "cures" do more harm than good, and others are simply ineffective. Fortunately, a host of interventions are effective. But this finding begs the question "How do we know?"

THE EVIDENCE-BASED APPROACH

We define the evidence-based approach as a conscientious effort to make evidence-based decisions regarding programs, practices, and larger-scale

policies to ameliorate antisocial behaviors (Petrosino & Boruch, 2014). The evidence-based approach has become "all the rage" (Tilley & Laycock, 2000) in the United States and elsewhere due to a number of factors, including the perceived need to channel scarce resources toward demonstrably effective treatments and to not support ineffective or harmful treatments. According to this approach, research evidence is viewed as a more appropriate basis for decision making regarding resource allocation than "common sense," tradition, ideology, and politics.

Still, one might ask, "How does one identify an effective treatment for antisocial behavior?" or determine if a treatment program is harmful? The evidence-based movement has focused considerable attention on (1) rigorous experimental and quasi-experimental impact evaluations, (2) syntheses of comparable studies through systematic reviews and meta-analyses, and (3) program ratings provided by evidence-based registries and best practice lists.

To identify programs for which the evidence points to harmful or positive outcomes, we relied on evidence from systematic reviews and meta-analyses of randomized controlled experiments and quasi-experimental evaluations, or through evidence-based registries such as Blueprints for Healthy Youth Development. Our search techniques were not systematic, as the goal of the chapter was to provide some examples in each group rather than to comprehensively summarize all of the many treatments for antisocial behavior.

Systematic reviews use explicit and rigorous methods to synthesize similar studies, and are designed to reduce the potential for bias that can occur when summarizing literature (Boruch & Petrosino, 2010). Meta-analyses apply quantitative methods to the synthesis process. Systematic reviews and meta-analyses relevant to the treatment of antisocial behavior of children were identified by searching the libraries of the international Campbell Collaboration (C2) and Cochrane Collaboration in health care, as well as those published in journals.

In addition to systematic reviews, another useful resource we relied on is the evidence-based registry, an electronic, online resource that identifies screens and rates the strength of evidence for a specific intervention's efficacy claims. Evidence-based registries and best practice lists provide ratings of how effective a specific, fine-grained intervention (such as functional family therapy, for example) is according to carefully screened evidence. We examined a number of evidence-based registries for identifying effective prevention programs—the U.S. Department of Justice's Crime Solutions, the U.S. Office of Juvenile Justice and Delinquency Prevention's Model Programs, the U.S. Substance Abuse and Mental Health Services Administration National Registry of Effective Policy and Practice (NREPP), the World Health Organization's Violence Prevention Evidence Base, and the University of Colorado's Blueprints for Healthy

Youth Development—to build on our understanding of the efficacy of the approaches we reviewed.

The advent of electronic publication and instant access to resources through the World Wide Web has meant that evidence from systematic reviews and evidence-based registries can be disseminated rapidly to busy decision makers. It also meant that we could quickly assemble them to identify examples of programs cited here.

Most prevention and intervention programs currently in place have not been adequately evaluated. This state of affairs is not unusual. For example, in a study of school-based drug prevention programs, Ringwalt and his colleagues (2011) reported that less than half (46.9%) of schools in their national sample used evidence-based programs. This was an improvement over a prior survey, indicating the growth of evidence-based curricula, but it indicates that more than half of the schools were using locally created and untested programs.

For this chapter, a harmful effect can be defined as one in which the review, meta-analysis, or evidence-based registry summarizes the evidence and indicates that the counterfactual—the control or comparison group—fared better than the treatment group on outcomes of crime and antisocial behavior. In contrast, an effective program is one in which the evidence indicates that the treatment group performs better than the counterfactual condition on such outcomes.

Because there is no readily accessible way to identify novel or innovative programs, we queried researchers and searched the World Wide Web for examples of strategies being used with antisocial children that have not been tested through rigorous impact evaluation to determine their effectiveness. Some of the treatments we list as novel therapies may be harmful using other criteria (e.g., attachment therapy has no rigorous evidence supporting claims of benefit or harm, but it may be harmful to its participants given the fatalities that have occurred due to coercive restraint; see Mercer, Chapter 15, this volume). Both systematic reviews and evidence-based registries that we searched focused on rigorous impact studies. We define a rigorous impact study as one that either randomly assigns participants or other units (including aggregate units such as communities or schools) to conditions, or uses quasi-experimental techniques such as matching or other methods for generating equitable comparison groups. This does not mean that all other evaluations are biased, but in determining the effectiveness of a program, experimental and quasi-experimental designs are more convincing in that they most effectively rule out alternative explanations for observed results (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002).

Although there are several types of research validity, including construct validity (i.e., the appropriateness and adequacy of the definitions of variables and measurements in the study) and statistical conclusion validity

(i.e., the trustworthiness of the results), most systematic reviews/meta-analyses and evidence-based registries emphasize internal validity. Internal validity refers to how well a study demonstrates that an intervention produced an effect on an outcome while ruling out potential rival causal factors. The establishment of internal validity typically requires a control or comparison group that allows the researcher to estimate what would have happened to the experimental group had the intervention *not* been applied (i.e., the "hypothetical counterfactual"). In short, a study with high internal validity is considered to have controlled for all plausible rival explanations for the observed results, except for participation in the treatment. A study with low internal validity, on the other hand, does not adequately control for rival hypotheses (Weisburd, Petrosino, & Fronius, 2014).

HARMFUL PSYCHOLOGICAL AND OTHER TREATMENTS

In this section, we review several harmful or ineffective treatments or responses to antisocial behavior of children. McCord (2003) summarized several "cures that harm," and there are numerous examples in the literature suggesting that even well-meaning interventions can backfire. As the examples we provided indicate, some are so popular that they have been featured uncritically on television shows (e.g., Scared Straight) and otherwise continue in the field today.

Individual Casework

A common-sense idea that first percolated in the 1930s was to assign wayward youth to a counselor. This counselor would oversee the child's development, including acting as an advocate and broker of services the child needed to move from antisocial behavior. This idea was based on the theory of individual deficits; antisocial youth had more deficits than law-abiding youth, and all that was needed was to address those deficiencies through some type of treatment.

One of the most famous examples in the criminological literature of the failure of this type of casework is the Cambridge–Somerville Youth Experiment (Jain & Cohen, 2014). This was a longitudinal study of 650 at-risk "difficult" boys that began in the 1930s in Cambridge and Somerville (Massachusetts) and is considered by many to be the first crime study that used randomization (Weisburd & Petrosino, 2005). The study categorized the 650 boys into 325 pairs based on over 100 matching characteristics; one member of each pair was randomized to treatment, with the other assigned to control. The treatment group was assigned to a "friendly counselor" and received a variety of services over a decade, including health, education, and coordination of community agencies and recreational services (Jain & Cohen, 2014). The control group received no special services.

Although the overall results were somewhat mixed (Jain & Cohen, 2014), the long-term effects on delinquency and crime were nonexistent or negative. McCord (1978) conducted the most extensive follow-ups and reported that the program had almost no positive effects on improving the lives of treatment participants on a number of outcomes, including committing additional crimes, health, and level of satisfaction with work. One major negative finding was that a higher percentage of treatment participants went on to commit additional crimes than control group members. Another negative finding was that the most harmful impact was found for youth who received the strongest dosage (Jain & Cohen, 2014; McCord, 2003).

How could such a benevolently constructed intervention have harmful effects? McCord spent considerable time trying to identify the mechanisms by which casework, as implemented in this study, did not work and may have even harmed. For example, there may have been a labeling effect, given that treatment boys were more visible as being difficult and at-risk because of their assignment to a counselor (control participants received nothing). In her later analyses, one potential mechanism for toxic effects identified by McCord was that constant attention and encouragement by the counselors raised expectations among boys receiving casework that could not be fulfilled in the lower-income, troubled neighborhoods and backgrounds of the at-risk youth (McCord, 2003). Another important mechanism was the role of peer contagion, or bringing deviant youth together to treat them in peer groups; this occurred because most youth in the treatment group attended summer camp together (McCord, 2003).

Peer Group Interventions

The Cambridge–Somerville Youth Experiment brought attention to the ineffective (and potentially harmful) use of peer group interventions. In the context of treatment of antisocial behavior, peer group interventions typically involve the use of discussion groups that bring together those children and youths with similar antisocial behavior diagnoses. Previous research indicated that peer group interventions involving deviant youth are not merely ineffective, but potentially harmful (Dishion, McCord, & Poulin, 1999). The two studies reviewed by Dishion and his colleagues (1999) included the Cambridge–Somerville Youth Experiment and the Adolescent Transitions Program (ATP). In the ATP study, 119 high-risk teens (both male and female) and their families were randomized to four conditions: (1) a parent focus group, (2) a teen focus group, (3) a parent and teen focus group, or (4) a placebo group (videotapes and written material of prosocial behavior). They attended the groups for 12 weeks. At 3-month outcomes, the teen focus group did worse, with increased teachers' reports of delinquent behaviors, and this persisted 3 years after the intervention.

These findings led Dishion et al. (1999) to describe the process by

which toxic effects emerge as "deviancy training." In other words, participating youth learn new modalities of antisocial behavior through observation and modeling processes during interactions with other antisocial peers during the discussion group meetings. A more recent meta-analysis by Weiss and colleagues (2005) examined 66 studies involving 115 treatment groups. They concluded that there was little evidence for consistent harmful effects of treatment in peer groups across these studies. However, they also reported that using antisocial youth in peer groups did not lead to positive impacts or reductions in antisocial behavior.

Juvenile Transfer Laws

Some antisocial youth, as in the Washington case we cited at the start of the chapter, commit very serious acts. Because most juvenile dispositions or punishments are only binding to when a child reaches the maximum age of jurisdiction of the juvenile court, youths committing particularly heinous offenses might only be held until age 21. Almost all jurisdictions in the United States have responded by providing a mechanism by which these serious juvenile offenders can be transferred to the jurisdiction of the adult court (the criminal justice system, as opposed to the juvenile justice system) and therefore tried as adults. Despite the notorious juvenile cases that often drive transfer policies, most crimes for which prosecutors or district attorneys seek a waiver are nonviolent in nature (Adams & Addie, 2010).

Besides the apparent net-widening that has occurred in many jurisdictions, Redding's (2010) broad review of the research found that the use of juvenile transfer to adult courts resulted in higher recidivism rates among this group compared with those juveniles who were not waived but instead were handled by juvenile courts. Redding examined six quasi-experimental studies that focused on whether transfer impacted subsequent offending by comparing juveniles transferred to adult court with youth who remained in the juvenile court, using matching or statistical controls to increase the similarity of the groups. All six studies indicated that transfer was associated with higher recidivism rates than for juveniles who were not transferred. Redding's (2010) review offers several explanations for the higher recidivism rates: a lack of focus on rehabilitation and family support and negative effects associated with being labeled by the adult court as a convicted felon.

Redding also pointed to a small study in which juvenile offenders were interviewed after being charged as adults. In the study conducted by Redding and Fuller (2004), a high percentage (78%) of juvenile offenders reported that they were unaware of the transfer laws in their state and that they expected to receive lighter sentences from juvenile courts (Redding, 2010). If juvenile transfer youths are meant to "deter" youth by scaring them with additional penalties, these limited data indicate the message is not coming through clearly.

Scared Straight and Other Juvenile Awareness Programs

Scared Straight and other juvenile awareness programs organize visits of juvenile delinquents or at-risk youth to prison facilities, usually adult institutions, with the aim of deterring the participants from future offending. The mechanism through which this behavior is deterred involves firsthand observation and experience of prison life as well as interactions with incarcerated adults. Petrosino, Turpin-Petrosino, and Finckenauer (2000) and Petrosino, Turpin-Petrosino, and Buehler (2005) conducted a systematic review on Scared Straight and related programs. They identified nine randomized experiments conducted in the United States from 1967 to 1992 that tested Scared Straight or similar juvenile awareness program. In the studies, youth were randomly assigned to attend either Scared Straight or a no-treatment control group. A meta-analysis across the studies indicated that the program had harmful effects; that is, the youth who attended Scared Straight did worse than youth who received nothing.

For example, in the most famous experiment, Finckenauer (1982) randomized youth to attend Scared Straight at Rahway State Prison or to a no-treatment control group. Youth in the treatment program were screamed at by inmates and warned about prison life, including brutal descriptions of rape and assault. In a 6-month follow-up, Finckenauer (1982) found that youth who attended Scared Straight were substantially more likely to be rearrested. The study propelled Finckenauer (1982) to write a book discouraging beliefs that a low-cost and easy-to-implement panacea for crime prevention and control, like Scared Straight, could be identified.

There have been several post hoc attempts to explain the apparent harmful effects of Scared Straight, including the explanation that the youth did not take the adult inmates' threats or warnings seriously (treating it like a "dare"). Another theory, discussed earlier, is that Scared Straight is delivered in peer groups who attend together. As a consequence, a peer contagion effect may have occurred by which law-breaking youth influence each other to get in trouble (McCord, 2003). Unfortunately, the data reported in the evaluation studies did not lend themselves to careful analysis to unpack the critical ingredients or mechanisms about why Scared Straight failed.

Juvenile Processing

Scared Straight and other similar programs imply deterrence, but it is a "threat" made by incarcerated inmates to the youth attending the program. In some sense, it is a third-party threat and may not be very compelling to antisocial youth. But what is the deterrent impact of the formal juvenile justice system?

There are several juvenile justice system stakeholders with discretionary decision-making authority over the response to youths following arrest, from the first contact with the police to the formal disposition (sentencing)

stage in the juvenile court. At any point, the decision makers can determine to divert the youth out of the formal juvenile justice system, either to be released to his or her parents or to participate in a diversion program with services such as counseling provided.

Petrosino, Turpin-Petrosino, and Guckenburg (2010) conducted a systematic review of research on the effects of juvenile system processing on subsequent delinquency. This systematic review included 29 experiments and over 7,300 juveniles, most of whom committed lower-level juvenile offenses such as trespassing, vandalism, shoplifting, underage drinking or smoking, and theft, but a few studies included more serious offenders (e.g., robbery). Results indicated that formal processing of juvenile offenders actually increases reoffending. The difference between processing and diversion alone is small, but in favor of diversion; essentially, there is a slight harmful impact to formally moving a juvenile through the court system when compared with doing nothing at all. A starker contrast was reported when comparing juveniles who were formally processed with those who were diverted to receive services. The authors concluded that these results indicate that formal processing not only may have a harmful effect (increasing delinquency) but society is actually paying more for it, as diversion—even with services—is usually cheaper than formally moving a youth through the system.

Again, as in the case of Scared Straight (Petrosino et al., 2005), data that would allow the authors to dismantle the mechanisms for why processing increases subsequent offending were not available. One theory is that the increase in recidivism for the processing group is due to labeling effects; that is, because the youth is officially processed, police, schools, and significant others now view him or her as a criminal. This labeling effect leads to the youth falling under more scrutiny and surveillance.

Because many youth commit acts that could meet legal definitions of delinquency, it is the increased surveillance—not any real bump in offending—that leads to the negative effect for processing. However, the authors examined the small number of studies that included self-reported outcomes (as opposed to official outcomes such as police arrest or court appearance). Even with self-reported data, formal processing seems to have backfired by increasing offenses.

Boot Camps

Another common-sense notion is that antisocial youth lack discipline and that all that is needed to turn them around is strict, militaristic regimentation. From those notions, boot camps emerged as a response to antisocial behavior of youth and young adults. In the typical boot camp, youth undergo rigorous physical conditioning and extreme disciplinary techniques, and follow strict rules modeled after military training approaches. Youth also take part in activities that are designed to improve participant

self-esteem, confidence, and leadership capacities. Many boot camp programs also incorporate other components such as physical labor, drug treatment, mental health treatment, and educational activities.

The typical boot camp program lasts between 90 and 180 days, and following the program the offender typically returns to the community. The intended mechanism for reducing recidivism involves breaking down the individual (the "self") and then rebuilding that person as a member of a group or team (with that group affiliation becoming the primary source of identity for the individual).

Wilson, Mackenzie, and Mitchell (2005) conducted a systematic review of the effects of boot camps on adults and juvenile offenders. They examined 32 studies and found little difference between treatment and control groups (Wilson et al., 2005). Nevertheless, there were substantial variations across studies in outcome; some evaluations indicated that the treatment group did better than the controls, whereas others showed the opposite. Wilson et al. (2005) found no conclusive differences for males or females sent to boot camp, and methodological features (type of study design), age of the offender, or type of crime the offender committed also seemed to have no effect on whether boot camp was effective or not. Wilson et al. (2005) did report that programs with a therapeutic component fared better (exhibited larger effects) than those without one, but there were too few studies to draw a definitive conclusion. Overall, Wilson et al. (2005) suggested that the military component of boot camps is not effective in reducing recidivism among offenders, as it is the common characteristic among all boot camps.

EFFECTIVE TREATMENTS

Interventions can exert a wide range of effects. The preceding section identified several treatments that appear to be either largely ineffective or harmful. Nevertheless, there are also some interventions that are demonstrably effective in reducing antisocial behavior of youth. This section highlights a few interventions that have yielded robust evidence of positive impact.

Multisystemic Therapy

Most seriously antisocial youth do not have a single "cause" for why they are committing criminal acts, but typically they confront a number of problems at home, school, individually, and with peers. Multisystemic therapy (MST) is a multimodal intervention that targets serious juvenile offenders by providing a comprehensive array of services to address these multiple areas of need (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998). That treatment can include a range of strategies,

including individual, family, peer-based, school, and community-based interventions. More often than not, MST is referred to as family-based treatment.

MST has been tested in a number of trials, and evidence has been synthesized in at least two meta-analyses. Henggeler, Clingempeel, Brondino, and Pickrel (2002) conducted a randomized experiment of MST, comparing youth receiving MST to those in a control group receiving usual community services. The MST treatment group had a significantly lower average rate of yearly convictions for aggressive criminal activity. The same did not hold true for property crimes, however. Findings with respect to long-term illicit drug use were mixed. Similarly, Schaeffer and Borduin (2005) compared MST with individual therapy using a randomized experimental design. The treatment group (MST) exhibited significantly lower recidivism rates than the control group, including lower rates of rearrest for violent offenses. Furthermore, the treatment group had 54% fewer arrests as well as 57% fewer days of confinement in adult detention facilities later on. Borduin, Schaeffer, and Heiblum (2009) compared a group randomly assigned to receive MST with a control group that received the usual community services for high-risk juvenile sex offenders. The MST group reported lower recidivism for sexual and nonsexual crimes. The treatment group also had 70% fewer arrests for all crimes and spent 80% fewer days in detention facilities compared with the control group. These studies are particularly informative as they provided positive results from long-term follow-ups that lasted into the early adult years.

Several systematic published reviews have examined MST; unfortunately, two of them conflict in their bottom-line conclusions. Following the line of results reported above, a systematic review by Curtis, Ronan, and Borduin (2004) indicated that MST reduced subsequent reoffending among serious (and sometimes violent) juveniles. In contrast, Littell, Campbell, Green, and Toews (2005) published a Campbell Collaboration review indicating that the findings for MST largely dissipated when more rigorous synthesis methods were applied. Lux, Jonson, and Cullen (2012) recently conducted a meta-analysis of fifty MST outcome studies and concluded that MST is an effective strategy for dealing with antisocial youth and their families. Collectively, it appears that the evidence lands on the side of average, positive impacts for MST, although questions regarding its long-term effects remain.

Functional Family Therapy

Given the role of poor parenting practices as discussed earlier as a risk factor, there is no surprise that a number of family-based interventions for responding to antisocial behavior of children would emerge. One of these interventions is functional family therapy (FFT). This treatment attempts to modify the patterns of family interaction to encourage clearer communication among parents and children and to minimize conflict between

members. The mechanisms through which this is accomplished include modeling, prompting, and reinforcement (Alexander & Parsons, 1973).

Gordon, Graves, and Arbuthnot (1995) examined FFT with rural, poor, white youth in Ohio using a quasi-experimental design with matching, although more severe cases were assigned to FFT. Twenty-seven juveniles were court ordered to attend FFT, and 27 comparison juveniles were believed to have received little or no counseling from mental health professionals. At first follow-up (28 months), Gordon et al. (1995) reported that only 11% of the FFT group reoffended compared to 67% in the regular services group.

Multidimensional Treatment Foster Care

Many youth in foster care and group home settings are harmed due to child abuse or other problems in their early childhood years. Some treatment programs have been designed to address these issues. For example, multidimensional treatment foster care (MTFC) uses individual-focused therapeutic care for adolescents who live in alternative, noncorrectional environments, primarily foster care. This approach also incorporates parent management training (Chamberlain & Reid, 1998).

Chamberlain, Leve, and DeGarmo (2007) carried out a randomized controlled trial comparing MTFC with a control group exposed to group care for serious and chronic female juvenile offenders. MTFC was more effective than group care (measured by days in locked settings, number of criminal referrals, and self-reported delinquency rates). Eddy, Bridges Whaley, and Chamberlain (2004) conducted an RCT of MTFC compared with a services-as-usual group home care control group for adolescent males who were under the control of the juvenile justice system. MTFC was significantly more effective than group home care (measured by referrals for violent offending and self-reported violent behavior). Moreover, rates of self-reported violent offending were 4 to 9 times higher among the group home care participants compared with those who received MTFC.

MacDonald and Turner (2007) conducted a systematic review of MTFC for the Cochrane Collaboration. Their stringent inclusion criteria meant that only five studies were eligible for the review. MacDonald and Turner (2007) concluded that "data suggest that treatment foster care may be a useful intervention for children and young people with complex emotional, psychological and behavioural need, who are at risk of placements in non-family settings that restrict their liberty and opportunities for social inclusion" (p. 8).

Cognitive-Behavioral Therapy

One of the risk factors for antisocial behavior is negative thinking patterns or cognitive distortions that fuel harmful acts. Cognitive-behavioral therapy (CBT) approaches were developed to help antisocial youth to identify and correct these problematic thoughts. CBT usually encompasses

several different techniques, such as social skills training, moral reasoning, aggression management, and other components to form a comprehensive approach to the offender's thinking patterns.

Landenberger and Lipsey (2005) conducted a meta-analysis of the effects of CBT on juvenile and adult offenders. They found 14 experimental or quasi-experimental studies to include in their analysis. They concluded that the overall results across these studies indicate that CBT reduces subsequent recidivism. Other meta-analytic results reported by Wilson, Bouffard, and Mackenzie (2005a); Pearson, Lipton, Cleland, and Yee (2002); Aos, Phipps, Barnoski, and Lieb (2001); and Tong and Farrington (2006) further support the efficacy of these programs in reducing antisocial and delinquent/offending behaviors.

NOVEL OR UNTESTED TREATMENTS FOR ANTISOCIAL BEHAVIOR

Given the harm to society that results from antisocial behavior by children, it is not a surprise that a range of strategies would be developed in an attempt to address it. Many or most of these strategies, however, have not been tested in impact evaluation. This does not mean the treatments are harmful, although in at least one case below, there is strong suggestive evidence that it may be (attachment therapy). It also does not mean the treatment is effective; the scientific evidence does not presently allow a firm conclusion.

Attachment Therapy

Another approach to treating antisocial behaviors and conduct disorder involves attachment therapies (see also Mercer, Chapter 15, this volume). These therapies include treatment methods such as "rebirthing," "reparenting," and "holding." These treatments focus on children with conduct disorder who are having difficulty in developing appropriate bonds with their parents. In some instances, children may have been separated from birth parents and never had an opportunity to bond with their biological parents or any other caregiver. Children who are adopted or are in the foster care system are supposedly at greater risk for attachment disorders (see Mercer, Chapter 15, this volume).

In rebirthing, the child reenacts or relives the experience of birth with the aid of his or her therapist. This type of therapy is sometimes used following "holding therapy." In holding therapy, the child or adolescent is held for extended periods of time by the current "parent" figure until a predetermined amount of time has passed or until the treated individual has stopped trying to break free of the embrace. The child then goes through a regressive treatment in which the parent provides direct care to the child in the form of hand and bottle feeding, wearing diapers that are changed by the parent, speaking "baby talk" to the child, and other similar strategies.

Some of these therapies use high levels of coercive, physical restraint and have been linked to several deaths by asphyxiation of the treated adolescents (see Mercer, Chapter 15, this volume). There is no high quality (or even moderate quality) evidence that such treatment modalities are effective in reducing conduct disorder problems or other antisocial behaviors (Mercer, 2002, 2003). Although a number of providers and organizations promote attachment therapy as a solution to the antisocial behavior of children who have not bonded with their parent or guardian, the psychological research community considers such treatment to be invalid (Chaffin et al., 2006).

Mentalizing-Based Therapy

Somewhat related to attachment therapy is mentalizing-based therapy (MBT). MBT has long been used as a treatment for patients with bipolar disorder and is now beginning to be used with antisocial persons. Undergirding MBT is the idea that the process of "mentalizing" or developing a sense of self, mindfulness, and control occurs in the early years of life, in the context of safe and secure child–caregiver relationships (Clarkin, Fonagy, & Gabbard, 2010, p. 328). However, mentalizing is often disrupted by abuse or neglect suffered during childhood (Clarkin et al., 2010, p. 328). MBT involves multiple interventions in which a therapist attempts to guide a patient to increased mentalization capacity, thereby leading to improved regulation of behavior and interpersonal relationships (Clarkin et al., 2010, p. 329). MBT has only recently been explored with adult patients, and it is too early to evaluate its effectiveness for adolescent patients. Future research will involve testing MBT's effectiveness in its various modalities (i.e., individual, group, and family settings), and across the spectrum of antisocial personality diagnoses.

Animal-Assisted Therapy

Given the positive response that interaction with certain animals provokes, a number of programs have surfaced that attempt to reduce anxiety and increase the comfort of children and adults in various contexts. Animal-assisted therapy (AAT) is often used in hospitals, assisted living facilities, and following community disasters to promote well-being. Not surprisingly, AAT is becoming an option for treating children and adolescents with antisocial and other behavioral problems. The rationale for what makes AAT therapeutic varies considerably, including such theories as being a source of social support (as in the case of the unconditional love of a dog toward its owner) or as an object of affection for the child.

Interestingly, animals have long been recognized by the mental health community for their ability to facilitate communication between two human parties, notably patients and therapists. Kruger, Trachtenberg, and Serpell (2004) suggested that the unique set of common and uncommon

characteristics that animals share with humans may generate positive social interaction in an otherwise uncomfortable or intimidating setting. Dogs are ideal candidates for positive social interactions because of their tolerant and unconditional positive regard for almost all humans. This interaction—brief or not—creates a level of comfort and calm for patients, especially in times of stress. This effect is relevant for adolescents who exhibit problem behaviors as the animal provides social support that may not be attainable by a human therapist. Animals also create natural bonds. These bonds encourage positive behaviors such as nurturing and caring because the animals rely on the human to take care of them (Kruger et al., 2004).

Besides dogs, dolphins are another animal that are used in this treatment; children who suffer from psychopathology, illness, or disability interact with dolphins in captivity, usually by swimming with them. Marino and Lilienfeld (2007) reviewed the evidence for "dolphin-assisted therapy" (DAT) and concluded that the studies were not sufficient to provide a definitive answer on the role of DAT in reducing antisocial behavior or other problematic behaviors in children and adolescents.

Although these examples highlight potential theoretical explanations for why AAT should work with adolescents, there is a lack of empirical evidence to support these theories. Future research should examine program efficacy through carefully done impact studies to allow for assessment of AAT.

Plastic Surgery for Juvenile Offenders

During the 1950s, institutions for juveniles and adult offenders began to offer surgery to correct physical deformities, especially those that were noticeable on the face. The rationale underlying such surgery was the belief that these physical deformities hindered the effective rehabilitation of the offender (Thompson, 1990). Another rationale for offering plastic surgery is the assumption that their antisocial conduct is due to their inability to "fit in" due to their disfigurement. Although there are published studies, we were unable to find one involving juveniles. In addition, most of the studies involving adults were plagued with internal validity problems. As Thompson (1990) noted, it is impossible from the literature to determine whether physical deformities were the cause of antisocial behavior—or a consequence of it (e.g., getting a mouth injury due to a knife wound).

IMPLICATIONS

As noted earlier, interventions can exert a range of effects, from a positive to a negative or harmful impact. How can the conscientious criminal justice or mental health professional discern what interventions to use? There are no guarantees, but it would seem prudent for practitioners and policymakers to hedge their bets and adopt programs that have demonstrated success in terms of reducing antisocial behavior of children.

Fortunately, the quest to identify evidence-based programs has been facilitated over the past two decades by methodological and technological developments (Baum, Beardslee, Lloyd, & Petrosino, 2013). On the methodological side, as mentioned earlier, a number of rigorous and explicit reviews of research on the effects of treatment programs for youth have been reported. Such reviews are often influential because they provide a comprehensive assessment of the body of evidence responding to a key policy question such as "What works?" Along with the systematic review, the evidence-based registry has facilitated the identification of very specific evidence-based treatment programs.

But a technological development, electronic publication through the World Wide Web, has made it possible for good evidence to be placed at the fingertips of most careful decision makers in a matter of seconds. Thus, the advent of systematic reviews and evidence-based registries, in concert with electronic publication, has meant that carefully vetted evidence is immediately available on a worldwide basis to decision makers concerned with implementing treatment programs. The concerned public servant or treatment professional would do well to consult reviews by the Campbell and Cochrane Collaborations, as well as the registries that were cited earlier in this chapter, to identify potentially effective treatments for the youth they serve.

But some challenges remain, as this survey of evidence on the treatment of antisocial behavior of children indicates. For one, despite the rise of evidence-based policy and the growth of scientific knowledge in this area, the number of reliable impact studies remains very small. Almost all systematic reviews and evidence-based registries sift through hundreds of study abstracts and usually end up with a small number of eligible studies. For example, Petrosino et al. (2005) examined over 500 abstracts on the Scared Straight program and included only nine in their meta-analysis. The number of studies being used to ascertain the effectiveness of other programs is even smaller. For example, the research on FFT is based on only four studies (Barnoski, 2002). Although the early results from these studies indicate positive findings, there is no guarantee that studies of the program in new settings will replicate the earlier evaluations.

Thus, more impact studies are needed across the board, not only of novel and untested treatments, but also of demonstrably effective programs that need to be tested in new settings or with different types of adolescent populations. This effort will ultimately enable more definitive and stable findings and conclusions about the effectiveness of treatment programs for troubled children and adolescents.

Another challenge to using research evidence to guide decisions about how to treat antisocial behavior of children is that harmful or ineffective programs are hard to remove from practice. This is not unusual; some people still smoke despite overwhelming medical evidence (physiological evidence of damage to lungs through X-rays and correlation research identifying smoking as a risk factor for later health issues). It is not a surprise

that when a popular program that is entrenched in the community is challenged by scientific evidence the program survives and evaluation research is ended. A classic example of this was the San Quentin prison version (known as SQUIRES) of the Scared Straight program. Lewis (1983) reported that youth in the control group in a randomized trial outperformed juveniles who went through the SQUIRES program. Finckenauer and Gavin (1999) wrote that when the Lewis (1983) evaluation came out, the response by the program was to get rid of the evaluators and use letters written to the prisoners as testimony of the program's success. Thus, evidence-based treatment for antisocial behavior of children and adolescents will not only require rigorous studies but a commitment on the part of decision makers to more intentionally use evidence when determining policy or practice.

GLOSSARY

Antisocial conduct: Pattern of behaviors that deviates from established social norms and intentionally causes harm to others or to society.

Antisocial personality disorder: A Mental health disorder diagnosed in individuals 18 years or older that is characterized by chronic and persistent displays of disregard for others and society's laws.

Conduct disorder: A Mental health disorder diagnosed in children and adolescents, generally before age 15, characterized by repetitive violations of others' rights and disregard for societal norms.

Evidence-based approach: Using evidence from previously conducted studies, usually systematic reviews and rigorous evaluation studies, to inform the selection of policies and programs to address a situation or problem.

Meta-analysis: A statistical method that analyzes data from documents reporting separate but similar experiments or studies to summarize results and to compare and contrast findings.

Randomized experiment: A type of evaluation study that uses random assignment, or chance probability, to assign units such as youth to treatment and control conditions. The randomized experiment is considered the most rigorous evaluation designs because it controls both known and unknown factors, other than participation in treatment.

Quasi-experiment: An experiment that lacks random assignment. Participants may be placed into groups based on a set of predetermined criteria (e.g., matched) or assigned in some other way to treatment and control groups without randomization.

Systematic review: A literature review of published studies that summarizes findings from different studies to try to answer a particular research question. Studies must fit a particular criterion established by the researcher in order to be included in the review.

ACKNOWLEDGMENTS

We appreciate the comments of the editors and Alison Cohen on prior drafts of this chapter and the recommendations about treatments from Todd Clear, Frank

Cullen, Denise Gottfredson, Frans Leeuw, Mark Lipsey, Kristin Moore, Haluk Soydan, and David Weisburd.

REFERENCES

Adams, B., & Addie, S. (2010). *Fact sheet: Delinquency cases waived to criminal court, 2007.* Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention

Alexander, J. F., & Parsons, B. V. (1973). Short-term behavioral intervention with delinquent families: Impact on family process and recidivism. *Journal of Abnormal Psychology*, *81*, 219–225.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Aos, S., Phipps, P., Barnoski, R., & Lieb, R. (2001). *The comparative costs and benefits of programs to reduce crime.* Olympia: Washington State Institute for Public Policy.

Barnoski, R. (2002). *Washington State's implementation of functional family therapy for juvenile offenders: Preliminary findings.* Olympia: Washington State Institute for Public Policy.

Baum, K., Beardslee, K., Lloyd, J., & Petrosino, A. (2013). *Violence prevention: Moving from evidence to implementation. Discussion paper.* Washington, DC: Institute of Medicine.

Borduin, C. M., Schaeffer, C. M., & Heiblum, N. (2009). A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: Effects on youth social ecology and criminal activity. *Journal of Consulting and Clinical Psychology*, *77*, 26–37.

Boruch, R. F., & Petrosino, A. (2010). Meta-analyses, systematic reviews and evaluation syntheses. In J. Wholey, H. Hatry, & K. Newcomer (Eds.), *The handbook of practical program evaluation* (3rd ed.). San Francisco: Jossey-Bass.

Chaffin, M., Hanson, R., Saunders, B. E., Nichols, T., Barnett, D., Zeanah, et al. (2006). Report of the APSAC Task Force on attachment therapy, reactive attachment disorder, and attachment problems. *Child Maltreatment, 11,* 76–89.

Chamberlain, P., & Reid, J. B. (1998). Comparison of two community alternatives to incarceration for chronic juvenile offenders. *Journal of Consulting and Clinical Psychology*, *66*, 624–633.

Chamberlain, P., Leve, L. D., & DeGarmo, D. S. (2007). Multidimensional treatment foster care for girls in the juvenile justice system: 2-year follow-up of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *75*, 187–193.

ChildStats.Gov. (2013). Child population: Number of children (in millions) ages 0–17 in the United States by age, 1950–2011 and projected 2012–2050. Retrieved March 1, 2013, from *www.childstats.gov/americaschildren/tables/pop1.asp.*

Clarkin, J., Fonagy, P., & Gabbard, G. (2010). *Psychodynamic psychotherapy for personality disorders.* Arlington, VA: American Psychiatric Publishing.

Cohen, M. A. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology, 19*, 75–78.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Crime and Disorder Act, 1998, C.37, UK Public General Acts. Retrieved July 10, 2012, from *www.legislation.gov.uk accessed*.

Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology, 18*(3), 411–419.

Dishion, T., McCord J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54*, 755–754.

Eddy, J. M., Bridges Whaley, R., & Chamberlain, P. (2004). The prevention of violent behavior by chronic and serious male juvenile offenders: A 2-year follow-up of a randomized clinical trial. *Journal of Emotional and Behavioral Disorders*, *12*, 2–8.

Eddy, J. M., & Reid, J. B. (2001). *The anti-social behavior of the adolescent children of incarcerated parents: A developmental perspective*. Eugene: Oregon Social Learning Center.

Eng, M. (2013). Prosecutor: Fifth-grade boys plotted to kill "annoying" girl. *MSN News*. Retrieved March 1, 2013, from *http://news.msn.com/us/prosecutor-fifth-grade-boys-plotted-to-kill-annoying-girl*.

Farrington, D. P., & Welsh, B. C. (2007). *Saving children from a life of crime: Early risk factors and effective interventions*. New York: Oxford University Press.

Finckenauer, J. O. (1982). *Scared straight and the panacea phenomenon*. Englewood Cliffs, NJ: Prentice-Hall.

Finckenauer, J. O., & Gavin, P. W. (1999). *Scared straight: The panacea phenomenon revisited*. Prospect Heights, IL: Waveland Press.

Gordon, D. A., Graves, K., & Arbuthnot, J. (1995). The effect of functional family therapy for delinquents on adult criminal behavior. *Criminal Justice and Behavior, 22*(1), 60–73.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Henggeler, S. W., Clingempeel, W. G., Brondino, M. J., & Pickrel, S. G. (2002). Four-year follow-up of multisystemic therapy with substance abusing and substance-dependent juvenile offenders. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 868–874.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. New York: Guilford Press.

Jain, S., & Cohen, A. (2014). Cambridge–Somerville Youth Experiment. In G. Bruinsma & D. Weisburd (Eds.), *Encylopedia of criminology and criminal justice*. New York: Springer.

Kazdin, A. E. (1995). *Conduct disorders in childhood and adolescence* (2nd ed.). Thousand Oaks, CA: Sage.

Kazdin, A. E. (1987). Treatment of antisocial behavior in children: Current status and future directions. *Psychological Bulletin*, *102*(2), 187–203.

Kazdin, A. E., Kraemer, H. C., Kessler, R, C., Kupfer, D. J., & Offord, D. R. (1997). Contributions of risk-factor research to developmental psychopathology. *Clinical Psychology Review*, *17*, 375–406.

Kruger, K. A., Trachtenberg, S. W., & Serpell, J. A. (2004). *Can animals help*

humans heal?: Animal-assisted interventions in adolescent mental health. Philadelphia: Center for the Interaction of Animals and Society, University of Pennsylvania School of Veterinary Medicine.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology, 1*, 451–476.

Lewis, R. V. (1983). Scared Straight—California style: Evaluation of the San Quentin *Squires* program. *Criminal Justice and Behavior, 10*(2), 209–225.

Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent and serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 86–105). Thousand Oaks, CA: Sage.

Littell, J., Campbell, M., Green, S., & Toews, B. (2005, October 29). Multisystemic therapy for social, emotional, and behavioral problems in youth aged 10–17. *Cochrane Database of Systematic Reviews*, *4*, CD004797.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., Moffitt, T. E., Caspi, A., White, H. R., et al. (2003). The development of male offending: Key findings from fourteen years of the Pittsburgh Youth Study. In T. P. Thornberry & M. D. Krohn (Eds.), *Taking stock of delinquency: An overview of findings from contemporary longitudinal studies* (pp. 93–136). New York: Kluwer/Plenum.

Lux, J., Jonson, C., & Cullen, F. (2012, November 14). *Assessing the effectiveness of multisystemic therapy: A meta-analysis*. Paper presented at the annual meeting of the ASC, Chicago, IL.

MacDonald, G. M., & Turner, W. (2007). Treatment foster care for improving outcomes in children and young people. *Campbell Systematic Reviews, 9*.

Marino, L., & Lilienfeld, S. O. (2007). Dolphin-assisted therapy: More flawed data, more flawed conclusions. *Anthrozoös, 20*(3), 239–249.

McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *Annals of the American Academy of Political and Social Science*, *587*, 16–30.

McCord, J. (1978). A thirty year follow-up of treatment effects. *American Psychologist, 33*(3), 284–289.

Mercer, J. (2002). Attachment therapy: A treatment without empirical support. *Scientific Review of Mental Health Practice,1*, 105–122.

Mercer, J. (2003). Violent therapies: The rationale behind a potentially harmful child psychotherapy. *Scientific Review of Mental Health Practice, 2*, 27–37.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*(4), 674–701.

Odgers, C. L., Milne, B., Caspi, A., Crump, R., Poulton, R., & Moffitt, T. E. (2007). Predicting prognosis for the conduct–problem boy: Can family history help? *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 1240–1249.

Pearson, F. S., Lipton, D. S., Cleland, C. M., & Yee, D. S. (2002). The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime and Delinquency*, *48*, 476–496.

Petrosino, A., & Boruch, R. F. (2014). Evidence-based policy in crime and justice.

G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 1437–1446). New York: Springer.

Petrosino, A., Derzon, J., & Lavenberg, J. (2009). The role of family in crime and delinquency: Evidence from prior quantitative reviews. *Southwest Journal of Criminal Justice*, *6*(2), 108–132.

Petrosino, A., Turpin-Petrosino, C., & Buehler, John. (2005). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency. *Scientific Review of Mental Health Practice, 4*(1), 48–54.

Petrosino, A., Turpin-Petrosino, C., & Finckenauer, J. O. (2000). Well-meaning programs can have harmful effects!: Lessons from experiments of programs such as Scared Straight. *Crime and Delinquency*, *46*(3), 354–380.

Petrosino, A., Turpin-Petrosino, C., & Guckenburg, S. (2010). Formal system processing of juveniles: Effects on delinquency. *Campbell Systematic Reviews, 1*.

Redding, R. E. (2010). *Juvenile Justice Bulletin: Juvenile transfer laws: An effective deterrent to delinquency?* Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Redding, R. E., & Fuller, E. J. (2004). What do juvenile offenders know about being tried as adults?: Implications for deterrence. *Juvenile and Family Court Journal*, *55*(3), 35–44.

Ringwalt, C., Vincus, V. A., Hanley, S., Ennett, S. T., Bowling, J. M., & Haws, S. (2011). The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2008. *Prevention Science, 12*(1), 63–69.

Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistance to psychiatric disorder. *British Journal of Psychiatry, 147*, 598–611.

Rutter, M., Giller, H., & Hagell, A. (1998). *Antisocial behavior by young people*. Cambridge, UK: Cambridge University Press.

Sampson, R. J., & Laub, J. H. (2005). A life-course view of the development of crime. *Annals of the American Academy of Political and Social Science*, *602*, 12–45, 73–79.

Schaeffer, C. M., & Borduin, C. M. (2005). Long-term follow-up to a randomized clinical trial of multisystemic therapy with serious and violent juvenile offenders. *Journal of Consulting and Clinical Psychology, 73*, 445–453.

Scott, S., Doolan, M., Beckett, C., Harry, S., Cartwright, S., & the HCA Team (2012). *How is parenting style related to child antisocial behavior?: Preliminary findings from the Helping Children Achieve Study*. London: Department of Education.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Singh, N. N., Lancioni, G. E., Singh Joy, S. D., Winton, A. W., Sabaawi, M., Wahler, R. G., et al. (2007). Adolescents with conduct disorder can be mindful of their aggressive behavior. *Journal of Emotional and Behavioral Disorders, 15*(1), 56–63.

Thompson, K. M. (1990). Refacing inmates: A critical appraisal of plastic surgery programs in prison. *Criminal Justice and Behavior, 17*(4), 448–466.

Tilley, N., & Laycock, G. (2000). Joining up research, policy and practice about crime. *Policy Studies, 2*(3), 213–227.

Tong, L. S. J., & Farrington, D. P. (2006). How effective is the "Reasoning and

Rehabilitation" programme in reducing reoffending?: A meta-analysis of evaluations in four countries. *Psychology, Crime and Law, 12*, 3–24.

Washburn, J. J., Romero, E., Welty, L. J., Abram, K. M., Teplin, L. A., McClelland, G. M., et al. (2007). Development of antisocial personality disorder in detained youths: The predictive value of mental disorders. *Journal of Consulting and Clinical Psychology*, *75*(2), 221–231.

Weisburd, D., & Petrosino, A. (2005). Randomized experiments. In J. Mitchell Miller & Richard Wright (Eds.), *Encyclopedia of criminology* (pp. 877–884). New York: Routledge.

Weisburd, D., Petrosino, A., & Fronius, T. (2014). Randomized experiments. *Encyclopedia of criminology and criminal justice*. New York: Springer.

Weiss, B., Caron, A., Ball, S., Tapp, J., Johnson, M., & Weisz, J. R. (2005). Iatrogenic effects of group treatment for antisocial youths. *Journal of Consulting and Clinical Psychology, 73*(6), 1036–1044.

Wilson, D. B., MacKenzie, D. L., & Mitchell, F. N. (2005). Effects of correctional boot camps on offending: A Campbell Collaboration systematic review. Available at *www.aic.gov.au/campbellcj/reviews/titles.html*.

CHAPTER SEVENTEEN

Science and Pseudoscience in Clinical Psychology

Concluding Thoughts and Constructive Remedies

Scott O. Lilienfeld, Steven Jay Lynn, and Jeffrey M. Lohr

We very much hope this edited volume, now in its second edition, has provided readers with at least a taste of the breadth and depth of the problem of pseudoscientific and otherwise questionable practices in clinical psychology and other mental professions, and offered helpful guideposts for distinguishing mental health claims with and without adequate empirical support. We also hope this volume has highlighted the continuing challenges posed by questionable assessment and intervention techniques in clinical psychology while highlighting the vast expanse of evidence-based interventions often available at the disposal of the contemporary practitioner. In our closing comments, we wish to propose several remedies to combat the continuing spread of potentially pseudoscientific claims in clinical psychology.

We believe the preceding chapters have made clear that the scientific underpinnings of the field of clinical psychology are threatened by the continuing proliferation of unsubstantiated and untested psychotherapeutic, assessment, and diagnostic techniques. Indeed, much of the book up to this point reads like a jeremiad. Readers who have made their way through most or all of the chapters in this book may be experiencing an understandable sense of pessimism, even despair, concerning the long-term future of clinical psychology and allied fields.

We believe such nihilism to be unwarranted. In this concluding chapter, we propose six remedies that we believe will go a substantial way toward healing the ills presently afflicting the field of clinical psychology. Our recommendations are a mixture of what Dawes (2005) termed hortatory and minatory standards, with the former being standards that exhort mental health professionals toward better practice ("carrots") and the latter being standards that threaten sanctions against those professionals who do not adhere to adequate practices ("sticks"). Despite the serious problems this book has highlighted, we are reasonably confident that if these six remedies are followed, the problem of pseudoscience in clinical psychology may ultimately prove amenable to a cure.

With no further ado, here is our six-point prescription for the field of clinical psychology:

1. All clinical psychology training programs must require formal training in critical thinking skills, particularly those needed to distinguish scientific from pseudoscientific methods of inquiry (e.g., see Gambrill, 2006, and Lilienfeld, Ammirati, & David, 2012). In particular, clinical training programs must emphasize such issues as (a) clinical judgment and prediction, and the factors (e.g., confirmation bias, hindsight bias, overconfidence, illusory correlation; Garb, 1998; see also Garb & Boyle, Chapter 2, this volume) that can lead clinicians astray when evaluating assessment information (see Grove, 2000, for similar recommendations); (b) fundamental issues in the philosophy of science, particularly the distinctions between scientific and nonscientific epistemologies (see Lilienfeld, Lynn, & Lohr, Chapter 1, this volume); (c) research methodologies required to evaluate the validity of assessment instruments (see Hunsley, Lee, Wood, & Taylor, Chapter 3, this volume) and the efficacy and effectiveness of psychotherapies (see Gaudiano, Dalrymple, Weinstock, & Lohr, Chapter 6, this volume); and (d) issues in the psychology of human memory, particularly the reconstructive nature of memory and the impact of suggestive therapeutic procedures on memory (see Lynn, Krackow, Loftus, Locke, & Lilienfeld, Chapter 8, this volume). Moreover, the American Psychological Association must be willing to withhold accreditation from clinical PhD and PsyD programs that do not place substantial emphasis on these and related topics, which in our view should be mandatory in the education and training of all clinical psychologists.

2. The field of clinical psychology must focus on identifying not only empirically supported treatments (ESTs; see Chambless & Ollendick, 2001), but also treatments that are clearly devoid of empirical support. In contrast to some critics (e.g., Westen, Novotny, & Thompson-Brenner, 2004) (see Gaudiano, Dalrymple, Weinstock, & Lohr, Chapter 6, this volume), we regard the effort to produce explicit lists of ESTs as laudable, although we share some of their thoughtful concerns regarding the

criteria used to identify these interventions. In particular, we believe that the criteria for ESTs need to be strengthened to include a heightened focus on (a) clinical, and not merely statistical, significance (Herbert, 2003), and (b) the theoretical plausibility of interventions (David & Montgomery, 2011).

The battle against pseudoscience is too substantial to be waged on only a single front. Although the identification of empirically supported therapies is an important long-term goal, we must also work toward identifying techniques that are either clearly inefficacious or harmful. The provisional list provided by Lilienfeld (2007; see also Dimidjian & Hollon, 2010), which includes crisis debriefing, facilitated communication, Scared Straight interventions, and certain attachment therapies, is a useful starting point, but it needs to be refined and updated in light of ongoing and future research.

3. The American Psychological Association and other psychological organizations must play a more active role in ensuring that the continuing education of practitioners is grounded in solid scientific evidence. Among the workshops for which the American Psychological Association has recently provided continuing education (CE) credits to practicing clinicians are courses in calligraphy therapy, Jungian sandplay therapy, and the use of psychological theater to "catalyze critical consciousness" (see Lilienfeld, 1998). The American Psychological Association has also offered CE credits for crisis debriefing, a technique that has been shown to be harmful in several controlled studies (see Lohr, Gist, Deacon, Devilly, & Varker, Chapter 10, this volume). Some state psychological associations have not done much better. The Minnesota Board of Psychology has even approved workshops in rock climbing, canoeing, sandplay therapy, and drumming meditation for CE credits (Lilienfeld, 1998).

If professional organizations intend to assist practitioners in the critical task of distinguishing techniques with and without adequate scientific support, they must insist on providing continuing education that serves this goal. Moreover, academics and clinicians who possess expertise in the differences between scientifically supported and unsupported assessment and therapeutic techniques must play a more active role in the development and dissemination of CE courses and workshops. To facilitate this process, academic clinical psychology programs must encourage their faculty members to participate in the construction and design of scientifically oriented CE courses.

4. The American Psychological Association and other psychological organizations must play a more visible public role in combating erroneous claims in the popular press and elsewhere (e.g., the Internet) regarding psychotherapeutic and assessment techniques. These organizations have traditionally been reluctant to play the role of media "watchdogs" in the battle against unsubstantiated mental health methods and claims. In an

era in which unsubstantiated mental health techniques are thriving with unabated vigor, however, such reluctance is becoming increasingly difficult to defend. The airwaves are increasingly dominated by talk show and media psychologists who dispense advice and information that are not supported by research evidence (Lilienfeld, 2012; Wilson, 2003), rather than by scientifically informed mental health professionals with the expertise necessary to provide the public with scientifically based information. As George Miller reminded us many years ago, "popular" psychology need not be a nonscientific psychology (Lilienfeld, 1998; Lilienfeld, Lynn, Ruscio, & Beyerstein, 2009).

We therefore strongly recommend that the American Psychological Association and other psychological organizations, including the Association for Psychological Science, create coordinated networks of media contacts (ideally consisting of experts who possess expertise regarding questionable or untested techniques in clinical psychology) who can respond to problematic or unsubstantiated mental health claims whenever they arise in the media, as well as to media inquiries regarding such claims.

5. The American Psychological Association and other psychological organizations must be willing to impose stiff sanctions on practitioners who engage in assessment and therapeutic practices that are not grounded in adequate science or that have been shown to be potentially harmful. The American Psychological Association Ethics Code clearly indicates that the use of unsubstantiated assessment techniques constitutes ethically inappropriate behavior. For example, American Psychological Association Ethics Code Rule 2.01(b) mandates that "Psychologists' assessments, recommendations, reports, and psychological diagnostic or evaluative statements are based on information and techniques (including personal interviews of the individual when appropriate) sufficient to provide appropriate substantiation for their findings" and American Psychological Association Ethics Code Rule 2.01(a) mandates that "psychologists do not base their assessment or intervention decisions or recommendations on data or test results that are outdated for the current purpose." The American Psychological Association Ethics Code (Rule 1.14) is similarly unambiguous in the case of potentially harmful psychotherapeutic methods: "Psychologists take reasonable steps to avoid harming their patients or clients, research participants, students, and others with whom they work, and to minimize harm where it is foreseeable and unavoidable."

Clinical psychologists who violate these codes of professional conduct must suffer appropriate consequences and must be prevented from harming the general public. Appropriate sanctions on the part of the APA and other professional organizations are a prerequisite for safeguarding the integrity of the profession and ensuring the safety of clients. *Primum non nocere.*

6. On a more positive note, the field of clinical psychology must actively address the continued sources of resistance to evidence-based

practice among many mental health professionals. Surveys indicate that large pluralities of clinical psychologists and other practitioners are dubious of the need for evidence-based practice, and believe that their clinical intuitions should supersede well-replicated controlled studies when they conflict with each other (Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013). Yet as a field, we have largely ignored the underlying sources of such resistance and have done little to address it. In particular, it will be essential for psychologists who are committed to a scientific approach to (a) proactively address potential concerns about evidence-based practice early on in the training of graduate students, and (b) work to establish partnerships between academics and science-oriented practitioners so that that the latter can more effectively address clinician resistances to evidence-based practice.

We modestly believe that this book should be required reading for all clinical psychologists and other mental health professionals, as well as all mental health professionals in training. We remain hopeful that if our six prescriptions are followed, a future edition of this book may be able to safely drop the words "and pseudoscience" from its title.

ACKNOWLEDGMENTS

We thank James Herbert and Richard McNally for their helpful comments on an earlier version of this chapter.

REFERENCES

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716.

David, D., & Montgomery, G. H. (2011). The scientific status of psychotherapies: A new evaluative framework for evidence-based psychosocial interventions. *Clinical Psychology*: *Science and Practice*, *18*, 89–99.

Dawes, R. M. (2005). The ethical implications of Paul Meehl's work on comparing clinical versus actuarial prediction methods. *Journal of Clinical Psychology*, *61*, 1245–1255.

Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist*, *65*, 21–33.

Gambrill, E. (2006). *Critical thinking in clinical practice*: *Improving the quality of judgments and decisions*. New York: Wiley.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.

Grove, W. M. (Chair). (2000). *APA Division 12 (Clinical) Presidential Task Force "Assessment for the year 2000": Report of the task force*. Washington, DC: American Psychological Association, Division 12 (Clinical Psychology).

Herbert, J. D. (2003). The science and practice of empirically supported treatments. *Behavior Modification*, *27*, 412–430.

Lilienfeld, S. O. (1998). Pseudoscience in contemporary clinical psychology: What it is and what we can do about it. *Clinical Psychologist*, *51*, 3–9.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*, 53–70.

Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist*, *67*, 111–129.

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *50*, 7–36.

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2009). *50 great myths of popular psychology: Shattering widespread misconceptions about human behavior.* New York: Wiley.

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review*, *33*, 883–900.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, *130*, 631–663.

Wilson, N. (2003). Commercializing mental health issues: Entertainment, advertising, and psychological advice. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 425–459). New York: Guilford Press.

Index

The letter *f* following a page number indicates figure; the letter *t* indicates table.

D

Q

R

S

T