

Perspectives on Psychological Science

<http://pps.sagepub.com/>

Why Ineffective Psychotherapies Appear to Work: A Taxonomy of Causes of Spurious Therapeutic Effectiveness

Scott O. Lilienfeld, Lorie A. Ritschel, Steven Jay Lynn, Robin L. Cautin and Robert D. Latzman
Perspectives on Psychological Science 2014 9: 355
DOI: 10.1177/1745691614535216

The online version of this article can be found at:
<http://pps.sagepub.com/content/9/4/355>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Association For Psychological Science](http://www.sagepub.com/content/9/4/355)

Additional services and information for *Perspectives on Psychological Science* can be found at:

Email Alerts: <http://pps.sagepub.com/cgi/alerts>

Subscriptions: <http://pps.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Why Ineffective Psychotherapies Appear to Work: A Taxonomy of Causes of Spurious Therapeutic Effectiveness

Scott O. Lilienfeld¹, Lorie A. Ritschel^{2,3}, Steven Jay Lynn⁴,
Robin L. Cautin⁵, and Robert D. Latzman⁶

¹Department of Psychology, Emory University; ²Department of Psychiatry, University of North Carolina at Chapel Hill; ³C Institute, Cary, NC; ⁴Department of Psychology, Binghamton University; ⁵Department of Psychology, Manhattanville College; and ⁶Department of Psychology, Georgia State University

Abstract

The past 40 years have generated numerous insights regarding errors in human reasoning. Arguably, clinical practice is the domain of applied psychology in which acknowledging and mitigating these errors is most crucial. We address one such set of errors here, namely, the tendency of some psychologists and other mental health professionals to assume that they can rely on informal clinical observations to infer whether treatments are effective. We delineate four broad, underlying cognitive impediments to accurately evaluating improvement in psychotherapy—naïve realism, confirmation bias, illusory causation, and the illusion of control. We then describe 26 *causes of spurious therapeutic effectiveness* (CSTEs), organized into a taxonomy of three overarching categories: (a) the perception of client change in its actual absence, (b) misinterpretations of actual client change stemming from extratherapeutic factors, and (c) misinterpretations of actual client change stemming from nonspecific treatment factors. These inferential errors can lead clinicians, clients, and researchers to misperceive useless or even harmful psychotherapies as effective. We (a) examine how methodological safeguards help to control for different CSTEs, (b) delineate fruitful directions for research on CSTEs, and (c) consider the implications of CSTEs for everyday clinical practice. An enhanced appreciation of the inferential problems posed by CSTEs may narrow the science–practice gap and foster a heightened appreciation of the need for the methodological safeguards afforded by evidence-based practice.

Keywords

psychotherapy, effectiveness, science–practice gap, confirmation bias, illusory correlation, spontaneous remission, placebo effect, regression to the mean, effectiveness, efficacy

A clinically depressed client obtains psychotherapy; 2 months later, she is free of serious symptoms. Was her improvement due to the treatment?

The correct answer is “We don’t know.” On the one hand, ample data demonstrate that scientifically supported psychotherapies can alleviate many mental health difficulties (Barlow, 2004), so the client’s improvement may well stem at least partly from the intervention. On the other hand, as most mental health professionals know, we cannot draw valid conclusions regarding a treatment’s effectiveness in the absence of methodological safeguards against errors in inference, such as well-validated outcome measures, randomized control groups, and blinded observations (Gambrell, 2012). Yet even seasoned clinicians and researchers can easily fall prey to

the error of concluding that a treatment worked when the evidence for this inference is insufficient. They can commit this mistake when evaluating the effectiveness of treatment for a given client, the effectiveness of a specific school or modality of psychotherapy, or both.

This error in reasoning can be found in published research as well. In numerous articles, authors have interpreted client improvement following an intervention—even in the absence of differences from a no-treatment control group—as evidence for treatment efficacy (e.g.,

Corresponding Author:

Scott O. Lilienfeld, Department of Psychology, Room 473, Emory University, 36 Eagle Row, Atlanta, GA 30322
E-mail: slilien@emory.edu

Leins et al., 2007). For example, in a study of psychological treatment for outpatients with severe depression, a research team randomized participants to receive either cognitive-behavioral or interpersonal therapy and found broadly equivalent improvement in both groups. Despite the absence of a no-treatment or placebo-control condition, the authors concluded that “both therapies are equally effective for depression” (Luty et al., 2007, p. 496; see also p. 500). More recently, in a randomized controlled study comparing psychoanalytic therapy with cognitive-behavioral therapy for bulimia nervosa—which also contained no control condition—the authors concluded that “Both treatments had substantial effects on global eating disorder psychopathology and general psychopathology” (Poulsen et al., 2014, p. 114).

In this article, we explain why the error of inferring that a treatment is effective on the basis of inadequate evidence is widespread, understandable, and problematic for clinical inference. We contend that a number of mental health professionals are insufficiently cognizant of the manifold reasons why ineffective or even harmful treatments can appear effective to the unaided eye. Because of this inadequate recognition, some clinicians and researchers may dismiss or minimize the need for evidence-based practice (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996; Straus et al., 2010).

Evidence-Based Practice and Causes of Spurious Therapeutic Effectiveness

Evidence-based practice is a threefold framework for clinical practice that is often conceptualized as a three-legged stool. These legs comprise (a) research findings regarding the efficacy and effectiveness of psychotherapies, (b) clinical expertise, and (c) client values and preferences (Norcross, Beutler, & Levant, 2007; Spring, 2007). Evidence-based practice is not synonymous with empirically supported therapies (ESTs), which are merely one set of operationalizations of the research leg of the evidence-based practice stool (Westen, Novotny, & Thompson-Brenner, 2005). ESTs are interventions that have been demonstrated to work better than no treatment (or an alternative treatment) for specific disorders in independently replicated (a) controlled between-subject designs or (b) controlled single-subject designs, namely, those in which participants serve as their own controls (Barlow, Hayes, & Nelson, 1984; Chambless & Hollon, 1998). Although the scientific status of ESTs is controversial (for diverse viewpoints, see Beutler, 2004; Castelnovo, 2010; Chambless & Ollendick, 2001; Herbert, 2003; and Westen, Novotny, & Thompson-Brenner, 2004), acceptance of the need for the research prong of evidence-based practice does not hinge on agreement with the criteria for or specific lists of ESTs, such as those proposed

by Division 12 (Society of Clinical Psychology) of the American Psychological Association (see <http://www.div12.org/empirically-supported-treatments/>).

The research leg, which is the component of evidence-based practice most pertinent to our arguments, incorporates control groups, within-subject designs, blinding, randomization, and other methodological bulwarks against inferential mistakes. In ways that have often not been adequately appreciated or articulated, these research safeguards are frequently nonintuitive. When viewed in this light, the much decried science–practice gap (Baker, McFall, & Shoham, 2008; Lilienfeld, Lynn, & Lohr, 2003; Tavris, 2003) and the resistance to evidence-based practice that often accompanies it (Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013) are not entirely surprising.

Although there are multiple sources of the science–practice gap (for discussions, see Lilienfeld et al., 2013; Shafran et al., 2009; Ritschel, 2005; and Stewart, Chambless, & Baron, 2011), we focus on one key contributor here: the myriad reasons why individuals can be led to conclude that psychotherapy is effective even when it is not. We term these sources of inferential error *causes of spurious therapeutic effectiveness* (CSTEs). Because of an insufficient recognition of CSTEs, psychologists may assume that they can rely on informal clinical observations of client change during and after treatment to gauge whether interventions are effective.

We do not contend that informal clinical observations of client improvement are never accurate; they frequently are. Nor do we argue that such observations are useless or should be disregarded, as they are at times helpful signposts of change in treatment. As noted earlier, substantial evidence attests to the efficacy and effectiveness of a broad swath of psychotherapies for many mental health conditions, including mood, anxiety, sleep, sexual, and eating disorders, as well as some personality disorders, such as borderline personality disorder (Roth & Fonagy, 2005; Wampold, 2001; Weisz, Weiss, Han, Granger, & Morton, 1995). Hence, clinicians' inferences of client improvement during and after psychological treatment are surely correct in many instances. Moreover, the social cognition literature demonstrates that numerous forms of intuitive thinking, such as heuristic processing, are often adaptive in real-world settings (Gigerenzer & Gaissmaier, 2011).

At the same time, the histories of medicine and psychology demonstrate that subjective inferences of change in treatment, subjectively compelling as they may be, are often mistaken (Garb, 1998; Grove & Meehl, 1996). Our overarching message is that because of CSTEs, unsystematic clinical observations of client change are rarely trustworthy guides by themselves for inferring treatment effectiveness.

Goals of the Article

Numerous articles have canvassed the magnitude and sources of the science–practice gap (e.g., Baker et al., 2008; McHugh & Barlow, 2010; Stewart et al., 2011). We do not intend to retread that well-traveled ground here. Instead, in light of relatively recent developments concerning (a) the implications of heuristics and biases for clinical practice (e.g., Crumlish & Kelly, 2009; Kahneman, 2011; Stanovich & West, 2008), (b) iatrogenic (i.e., psychologically harmful) effects in psychotherapy (e.g., Bootzin & Bailey, 2005; Dimidjian & Hollon, 2010; Lilienfeld, 2007), and (c) challenges to the dissemination of evidence-based practice (Lilienfeld et al., 2013; Stewart et al., 2011), we address the more specific and largely neglected question of what kinds of inferential errors in psychotherapy render evidence-based practice imperative.

The movement toward evidence-based practice has been contentious in many quarters, in part because some authors have taken issue with the premise that evidence derived from randomized controlled trials, controlled single-subject experiments, and other systematic research designs should be accorded higher priority than clinical experience when selecting treatments. Indeed, some scholars have proposed that “practice-based evidence,” namely, therapeutic practice informed by thoughtful clinical observations, should be accorded roughly equal weight to traditional evidence-based practice (Barkham, Hardy, & Mellor-Clark, 2010; Green & Latchford, 2012; Stricker, 2003). For example, Chwalisz (2003) lobbied for expanding the definition of evidence to encompass clinical observations and clinical consensus (see also Hoshmand & Polkinghorne, 1992). Similarly, while acknowledging that “practical knowledge” (viz., knowledge acquired from clinical observations of what does and does not work in treatment) is fallible, Bohart (2005) maintained that it is “evidence-based” (p. 46) and should be valued as a legitimate source of inferences for therapeutic effectiveness.

We view these assertions with decided ambivalence. On the one hand, clinical observations can sometimes (a) usefully guide therapists’ choices of interventions during treatment, (b) serve as springboards for the development of new models of treatment, and (c) inform the feasibility and transportability of scientifically based interventions to real-world settings. On the other hand, for reasons that we will explicate, the proposition that practice-based observations should be accorded comparable weight to the results of controlled clinical trials in treatment selection underestimates the inferential dangers stemming from CSTE. With this background in mind, our goals are threefold. First, we aim to demonstrate that inadequate appreciation of the inferential threats posed by CSTE is partly a by-product of natural cognitive processes that render it difficult for clinicians, clients, and researchers to

accurately perceive and evaluate therapeutic change. In addition, we delineate four broad obstacles to scientific thinking—naïve realism, confirmation bias, illusory causation, and the illusion of control—that underpin many or most CSTE. The distinction between these overarching cognitive impediments and specific CSTE themselves may not be entirely clear-cut. Nevertheless, we posit that these domain-general impediments lay the cognitive groundwork for more specific errors in inferring the existence or meaning of changes in treatment.

Second, we present a taxonomy of 26 CSTE, divided into three categories, that can contribute to the appearance of therapeutic effectiveness in its objective absence. These three classes of CSTE comprise influences that generate (a) the perception of client change in its actual absence, (b) misinterpretations of actual client change stemming from extratherapeutic factors, and (c) misinterpretations of actual client change stemming from non-specific treatment factors. Some CSTE operate at the level of individual clients, others at the level of groups of clients, and still others at both levels. Several writers in the medical literature have provided partial lists of artifacts that can make ineffective medical treatments seem effective (e.g., Beyerstein, 1997; Hall, 2011; Hartman, 2009; Kienle & Kiene, 1997), but no comparable list exists for psychotherapies; nor have previous authors provided a taxonomy of these artifacts.

Third, we outline how specific research methods help to control for, although not necessarily eliminate, CSTE as sources of erroneous conclusions in treatment. Although these research methods are by no means new, to our knowledge their role in helping to rule out differing CSTE has not been explicitly articulated. We also discuss how certain CSTE continue to pose unresolved challenges to psychotherapy researchers and point to fruitful areas for further research on CSTE and methods for attenuating their influence. In this respect, our analysis has heuristic value in that it points to gaps to be filled in extant psychotherapy methodology to minimize CSTE. Just as important, we discuss how a better appreciation of CSTE can inform everyday clinical practice. By promoting thoughtful consideration of explanations for client change above and beyond improvement due to therapy itself, CSTE can assist clinicians with becoming better clinical scientists. Finally, we demonstrate that our discussion of CSTE bears important implications for models of clinical training.

Overarching Cognitive Impediments to Evaluating Therapeutic Change

We submit that the principal reason why some mental health professionals may not appreciate sufficiently the problems posed by CSTE is that scientific thinking does

not come naturally to the human species (McCauley, 2011; Wolpert, 1992). Such thinking must be learned and practiced assiduously, because it often requires us to question our commonsense intuitions, such as our propensity to perceive meaningful causal relations in their absence (Bloom & Weisberg, 2007; Cromer, 1993). Most errors in judgment arising from CSTE probably reflect rapid and intuitively plausible perceptions and interpretations that do not sufficiently consider alternative explanations of client change.

One telltale sign of the counterintuitive nature of scientific thinking is the history of the concept of the control group. Contrary to what many psychologists might assume, this concept is a relatively recent development in scientific history (Bull, 1959; Dehue, 2000; Gehan & Lemak, 1994). Examples of controlled trials of medication surfaced only as recently as the 18th century, with the first arguably conducted by James Lind, who in 1747 famously divided sailors with scurvy onboard a British ship into several groups, finding that only those who received citrus juice improved (Manzi, 2012). Yet even Lind's discovery was apparently resisted, as the British Navy waited a full half-century before stocking lemon juice on its vessels (Bull, 1959). It was not until Coover and Angell (1907) advocated for the role of untreated groups in evaluating training programs in educational psychology that a formal exposition of the control group concept in social science appeared in print. The notion of the randomized control group is even more recent, emerging in the published literature in the 1920s (Dehue, 2005). Moreover, it was not until the 1950s that prominent authors (e.g., Eysenck, 1952; Meehl, 1955) began to call for randomized controlled trials of psychotherapy (Cautin, 2008).

As noted earlier, we contend that four broad cognitive impediments underlie many or most CSTE, which we view as specific instantiations of these impediments in the context of psychotherapy. We next lay out these impediments, followed by a brief discussion of their implications for therapists' perceptions of effectiveness.

Naive realism

Naive realism (Ross & Ward, 1996; Segall, Campbell, & Herskovits, 1966) is a concept imported into psychology from philosophy. Also termed *commonsense realism* or *direct realism*, naive realism is the ubiquitous assumption that the world is precisely as we see it. A plethora of phrases in everyday life attest to the power of naive realism in our thinking: "Seeing is believing," "I saw it with my own eyes," "I'll believe it when I see it," and "What you see is what you get." In a related vein, Kahneman (2011) referred to a core principle of intuition as "WYSIATI": What You See Is All There Is, an assumption

that dovetails with naive realism. This heuristic (mental shortcut) leads us to focus on what is most obvious in our environments while ignoring subtler background information.

Naive realism is erroneous because the world is not exactly as we perceive it, a point illustrated vividly by visual illusions (Chabris & Simons, 2010) and enshrined in the time-honored psychological distinction between sensation and perception (Coren, 2003). What we perceive is constrained by external reality, but it is also influenced by our expectations, biases, and interpretations ("apperceptions"; Morgan & Murray, 1935). To a substantial extent, "believing is seeing" as much as the converse (Gilovich, 1991).

Naive realism bears important implications for the evaluation of psychotherapy outcome. It can lead clinicians, researchers, and others to assume that they can rely on their intuitive judgments—"I saw the change with my own eyes"—to infer that an intervention is effective (Ghaemi, 2009; Lilienfeld, Lohr, & Olatunji, 2008). As a consequence, these individuals may (a) misperceive change when it does not occur, (b) misinterpret it when it does, or (c) both.

One example of the overreliance on naive realism comes from Arnold Shapiro, producer of the 1978 Academy Award-winning documentary *Scared Straight!*, who responded to scientific criticisms of Scared Straight interventions. These interventions attempt to frighten adolescents at high risk for crime out of criminal careers by bringing them to prisons and introducing them to inmates. Shapiro defended Scared Straight programs by insisting that "I'm seeing it [the change following Scared Straight programs] with my own eyes, I'm there for every one of those shoots" (Harrison, 2011, p. 2). However, data from controlled studies suggest that Scared Straight is not merely ineffective but probably harmful, in that it produces a heightened risk for antisocial behavior (Petrosino, Turpin-Petrosino, & Buehler, 2005). In another example, Healy (2002) wrote in an article, subtitled "Evidence-Based Psychiatry," that "When treatments work, the condition being treated vanishes, and we don't need randomized controlled trials to see this happening" (p. 1). Yet the condition being treated may disappear for a plethora of reasons other than the intervention. Contra Healy's implication, randomized controlled trials and other rigorous designs are indeed needed to exclude rival hypotheses for observed change.

Naive realism also reminds us of an easily forgotten principle: Change following therapy is not equivalent to change because of therapy, a logical error known as the *post hoc, ergo propter hoc* (after this, therefore because of this) fallacy (Finocchiaro, 1981). Conversely, this error can also lead individuals to equate deterioration following a treatment with deterioration because of the

treatment (Lilienfeld, 2007), thereby overestimating the iatrogenic effects of certain interventions. The *post hoc, ergo propter hoc* fallacy underscores the point that pre-post studies of interventions are problematic (T. D. Wilson, 2011). Fortunately, as we will discover, there are multiple ways to compensate for the limitations of pre-post designs. Investigations using such designs are especially suspect when the “pre” data derive from retrospective assessments. For example, the much ballyhooed *Consumer Reports* study (Seligman, 1995) of 4,100 magazine subscribers who had participated in psychotherapy revealed that most felt that they had been helped by it. Yet, as numerous critics (e.g., Jacobson & Christensen, 1996; Mintz, Drake, & Crits-Christoph, 1996) pointed out, these data are difficult to interpret, because the study neglected to control for many potential confounds that may have led to improvement even without therapy.

The history of medicine offers a powerful cautionary tale regarding the hazards of naive realism (Bigby, 1998). Most medical scholars agree that the history of physical treatments administered prior to about 1890 is essentially tantamount to the history of the placebo effect. Along with ineffective medications, such interventions as blood-letting, blistering, purging, and leeching were routinely prescribed and presumed to be beneficial based on little more than informal clinical observations (Grove & Meehl, 1996; see Belofsky, 2013, for a review of bizarre but widely accepted medical practices through the ages). Even today, medicine has its share of ineffective interventions. A recent meta-analysis estimated that 40% of widely used medical procedures (e.g., intensive glucose lowering in Type 2 diabetes, induction of hypothermia for intracranial aneurysms) are useless or harmful (Prasad et al., 2012).

The history of psychiatry is similarly replete with a litany of useless or harmful interventions, many of which were endorsed by experts of the era yet that strike us as inhumane today. Such “treatments” as spinning chairs, tranquilizing chairs, and cold water were ubiquitous in early American psychiatry. As another example, insulin coma therapy, introduced by Manfred Sakel in 1933, was used widely to treat schizophrenia throughout the 1930s and 1940s. This procedure involved administering increasingly high doses of insulin to induce a hypoglycemic state, followed by a coma and sometimes convulsions. Early clinical reports described encouraging results. Its high morbidity and mortality rates notwithstanding, insulin therapy spread rapidly throughout Europe, the United States, Japan, and Australia (James, 1992). This wave was unceremoniously interrupted by an article in the *Lancet* by Bourne (1953), who concluded there was no evidence that insulin coma therapy was effective. As Jones (2000) noted, many psychiatrists published

rebuttals to Bourne’s article: “Their tone was typified by remarks such as ‘it is clinical experience that counts here, despite all figures to the contrary’” (p. 148). By the late 1950s, insulin coma therapy had been all but abandoned (Shapiro & Shapiro, 1997).

Prefrontal lobotomy, which earned its principal developer of the procedure in humans, Portuguese neurosurgeon Egas Moniz, the Nobel Prize in Medicine or Physiology in 1949, offers another telling example. One practitioner of this technique insisted that “I am a sensitive observer, and my conclusion is that a vast majority of my patients get better as opposed to worse after my treatment” (see Dawes, 1994, p. 48), a view echoed by many of his contemporaries (Diefenbach, Diefenbach, Baumeister, & West, 1999). Later research, however, revealed lobotomy to be essentially worthless and to be associated with many disastrous psychological and neurological side effects (Valenstein, 1986).

Confirmation bias

A second cognitive impediment to appreciating the need for controls in psychotherapy research is confirmation bias. Confirmation bias is the deeply ingrained and commonly exercised tendency to seek out evidence consistent with one’s hypotheses and to deny, dismiss, or distort evidence that is not (Lilienfeld, Ammirati, & Landfield, 2009; Nickerson, 1998). Although confirmation bias is a cognitive phenomenon, it can be fueled by desires to find supportive evidence for our beliefs, a propensity termed “motivated reasoning” (Kunda, 1990). Because clinicians want their clients to improve, they can be driven to perceive change in its absence.

Confirmation bias can foster a propensity toward *illusory correlation* (not to be confused with illusory causation; see next section), which is the perception of a statistical association in its absence (Chapman & Chapman, 1967; Hamilton & Gifford, 1976). Specifically, confirmation bias can predispose clinicians to attend to the “hits” and forget the “misses” (Garb, Lilienfeld, & Fowler, 2008; Gilovich, 1991) and thereby overestimate the extent to which their interventions are associated with subsequent improvement. Imagine a therapist who engages from time to time in confrontational tactics with a client. Even though these tactics are ineffective for his client’s presenting problem, the therapist may attend to and recall the sessions in which the client was doing better and neglect and forget the sessions in which the client was not doing better or doing worse. As a consequence, the therapist may conclude that his use of confrontation was consistently followed by client improvement, even though it was not. In contrast, if the therapist were to monitor his clients’ symptoms systematically, this erroneous inference would presumably be less likely.

Illusory causation

Scottish philosopher David Hume (1748) maintained that humans are prone to perceiving causal relations in their absence. Two centuries later, Michotte (1945) argued that our propensity to perceive causal relations between events, even those that are causally unrelated, comes to us as naturally as does our propensity to perceive color. Research on *illusory causation*, or the propensity to perceive a spurious causal relation between two associated variables, bears out these contentions.

Laboratory evidence for illusory causation dates at least to the work of Koffka (1935), who showed observers two points of light in a dark room. When the points moved apart, perceivers tended to attribute causality to the dot on which they happened to be focusing, even if it was stationary. Koffka's findings suggest that we are more likely to attach causal significance to the object of our attention while ignoring competing evidence. Later research demonstrated that illusory causation extends to social interactions. When observers are positioned physically so as to attend primarily to one partner in a two-person conversation, they regard him or her as more interpersonally influential than the other partner (Taylor & Fiske, 1975; see also McArthur & Solomon, 1978).

There are two potential, nonmutually exclusive explanations for illusory causation (McArthur, 1980). The first is perceptual: Individuals tend to attribute causality to whatever stimulus is most vivid and prominent in their visual fields and to accord less causal import to what lies in the visual background (Lassiter, Geers, Munhall, Ploutz-Snyder, & Breitenbecher, 2002). The second is cognitive: Individuals recall more information about stimuli that are prominent in their visual foregrounds than in their visual backgrounds (Taylor & Fiske, 1978). With the aid of an availability heuristic, by which we gauge the probability of an event by using its accessibility in memory (Tversky & Kahneman, 1974), we come to view the former stimuli as more influential.

Because of illusory causation, therapists, researchers, clients, and external observers may leap to the conclusion that a treatment exerted a causal effect on the client when it did not (Sloman, 2005). The client's improvement within therapy sessions is plainly visible to the clinician, whereas rival explanations for this improvement (e.g., events occurring to the client outside of sessions, placebo effects, changes in cognitive biases over the course of treatment) rarely are. As a consequence, these explanations may be assigned less weight.

Illusion of control

A related error is the *illusion of control*, or the propensity to overestimate our ability to influence events (Langer,

1975). For example, when money is at stake, most people prefer to select a lottery ticket or roll a die themselves rather than leave these actions to others, even though the outcomes in all scenarios do not exceed chance. This illusion may predispose therapists to believe that they possess more causal power over client outcomes than they do. The illusion of control is especially likely when the individual in question (a) is personally involved in the behaviors, (b) is familiar with the situation at hand, (c) is aware of the desired outcome, and (d) has a history of previous success at the task (Thompson, 1999). Most or all of these criteria presumably apply to the modal psychotherapist. Indeed, when interventions are consistently followed by improvement, treatment providers may conclude that they are the active causal agents when they are not (Matute, Yarritu, & Vadillo, 2011).

Implications of cognitive impediments for clinicians' self-perceptions and predictions

These four broad cognitive impediments may help to explain why some therapists overestimate their positive client outcomes. In this respect, they appear to be no different from professionals in many other fields, including college professors (Cross, 1977), physicians (Hodges, Regehr, & Martin, 2001), and political pundits (Tetlock, 2005), all of whom tend to hold an overly charitable view of their effectiveness (Dunning, Heath, & Suls, 2004). In a sample of 129 therapists in private practice (26.4% psychologists), the average clinician rated him- or herself at the 80th percentile of all therapists in terms of effectiveness and skills; 25% of respondents placed themselves at the 90th percentile. None rated themselves as below average. Moreover, the typical therapist in the sample estimated the rate of client deterioration in his or her caseload to be 3.7% (Walfish, McAlister, O'Donnell, & Lambert, 2012). In fact, numerous studies have indicated that about 10% of clients become worse following psychotherapy (Boisvert & Faust, 2002; Lilienfeld, 2007).

Other evidence dovetails with these results. In a sample of 49 psychotherapists in college counseling centers, clinicians markedly overestimated their rates of positive client outcomes (91%) relative to their actual positive outcomes (40%), as ascertained by a standardized symptom measure. Furthermore, although therapists predicted that only 3 out of a total of 550 clients (0.5%) in their collective caseloads would deteriorate, outcome data revealed that 40 (7.3%) did so (Hannan et al., 2005). Taken together, these findings suggest that many or most psychotherapists perceive improvements in clients in their absence and fail to perceive deterioration in their presence.

Summary

In summary, four overarching cognitive biases—naïve realism, confirmation bias, illusory causation, and illusion of control—probably contribute to the difficulty of accurately evaluating change in psychotherapy, as well as to an insufficient appreciation of the inferential difficulties posed by CSTE, which we view as more specific instantiations of these four broad biases within the domain of psychotherapy. These broad biases may also contribute to clinician overconfidence, inadvertent neglect of adverse client outcomes, and an undue reliance on unguided clinical experience (see also Groopman, 2007).

Causes of Spurious Therapeutic Effectiveness: A List and Taxonomy

As noted earlier, we refer to the manifold ways in which people can be misled into believing that a treatment is working when it is not as *causes of spurious therapeutic effectiveness* (CSTEs). We next briefly describe 26 CSTEs that can deceive individuals into concluding that ineffective or even harmful psychotherapies are effective. We regard this list of CSTEs as provisional and subject to improvement pending further research. Hence, for heuristic purposes, we adopt a “splitting” rather than a “lumping” approach (see Mayr, 1981, for a discussion of the splitting–lumping dichotomy in classification) toward CSTEs, electing to subdivide them into distinct categories when there is research support for doing so. The advantage of a splitting approach is that certain CSTEs can later be combined into broader categories if evidence demonstrates that they are merely variants of the same inferential error.

In distinguishing among CSTEs, we part company with authors who have placed most or all CSTEs under the overarching rubric of placebo effects (e.g., Offit, 2010; Shapiro & Shapiro, 1997). For example, Novella (2008, 2010) defined placebo effects as “including everything other than a physiological response to a biologically active treatment” (p. 33) and operationalized it as “the treatment effect measured in the placebo arm of a clinical trial” (p. 33). Under placebo effects, Novella included such artifacts as regression to the mean, observer biases, demand characteristics, and expectancy effects. There are two shortcomings with this expansive conceptualization. First, it conflates the response following a placebo (the placebo response) with the response to a placebo (the placebo effect) and thereby runs afoul of the *post hoc, ergo propter hoc* error (Ernst & Resch, 1995; Kirsch, 2013). Many of the symptomatic changes that occur in a study’s placebo arm can arise from variables other than the placebo itself. Second, this conceptualization is overly

inclusive, because it does not distinguish among a myriad of sources of erroneous therapeutic effectiveness.

We stake no claim to our list’s comprehensiveness, but it provides a helpful starting point for conceptualizing the numerous challenges that confront clinicians, researchers, and clients when gauging psychotherapeutic effectiveness. Although all of the CSTEs we describe have been the subject of research on perceptions of change following interventions or experimental manipulations, several of these CSTEs (e.g., response-shift bias) have not, to our knowledge, been investigated with respect to psychotherapy per se. Nevertheless, there is no a priori reason why these latter CSTEs cannot produce the illusion of change following psychological treatment as well.

Overview of the taxonomy of CSTEs

We divide our proposed CSTEs into three overarching categories (see Table 1). The distinctions between these categories are conceptual, not empirical. First, some CSTEs, which we term Category 1 CSTEs, can lead individuals, including clinicians, researchers, and other observers, to misperceive change in its actual absence. In these cases, clients are not changing, although individuals erroneously perceive them to be changing. The problem of Category 1 CSTEs is underscored by a recent quotation from eminent psychiatrist Robert Spitzer, who 9 years earlier (Spitzer, 2003) had endorsed the effectiveness of “conversion therapies” for homosexuality on the basis of self-reported improvement from clients. In a widely publicized retraction of his conclusions, Spitzer (2012) acknowledged that there was no way to determine whether these perceptions of change were accurate. As Spitzer told a reporter (Carey, 2012, p. B1), “I knew this was a problem, a big problem, and one I couldn’t answer. How do you know someone has really changed?”

Category 1 CSTEs are highly heterogeneous, as some (e.g., CSTE Numbers 1 through 4; see following section) probably exert their initial effects primarily on clients’ perception of change, whereas others (e.g., CSTE Numbers 7 through 9 and 11) probably exert their initial effects primarily on clinicians’ perceptions of change. Still others (e.g., CSTE Numbers 10 and 13) probably exert their initial effects on both clients’ and clinicians’ perceptions. Nevertheless, because psychotherapy is a process of bidirectional influence between client and clinician (Marmar, 1990), most or all Category 1 CSTEs can eventually come to deceive both treatment recipient and treatment provider. Hence, these distinctions are unlikely to be clear-cut, and they will require empirical corroboration and potential revision.

Category 2 CSTEs can lead individuals, in most cases both clinicians and clients, to misattribute actual client

Table 1. Causes of Spurious Therapeutic Effectiveness and Research Safeguards Against Them

CSTE	Description	Research safeguards
Category 1 CSTEs: Erroneous perceptions of client change in its absence		All Category 1 CSTEs: Well-validated outcome indicators
1. Illusory placebo effects	Perceived improvement occurring in the absence of genuine improvement	(none additional)
2. Palliative benefits	Feeling better about one's signs and symptoms without tangible improvements in them	(none additional)
3. Confusing insight with improvement	Mistaking apparent understanding of one's problem with improvement in that problem	(none additional)
4. Retrospective rewriting of pre treatment functioning	Belief that one has improved arising from a tendency to remember one's pretreatment functioning as worse than it was	Measures of pretreatment functioning
5. Response-shift bias	Change in one's evaluation standard with respect to an outcome dimension as a consequence of treatment	Measures of pretreatment functioning
6. Reduction in cognitive biases	Declines in cognitive biases tied to pretreatment reporting of symptoms	Measures of cognitive biases throughout treatment
7. Demand characteristics	Tendency of clients to report improvement in accord with what they believe to be the therapist's or researcher's hypotheses	Outcome measures low in reactivity
8. The therapist's office error	Confusion of client's in-session behavioral presentation with out-of-session improvement	Out-of-session collateral reports of improvement
9. Test-retest artifacts	Tendency of scores on psychopathology measures to decline spuriously on their second administration	Outcome measures that do not contain a skip-out structure
10. Unknowable outcomes in the control condition	Lack of information regarding what would have occurred had the treatment not been administered	Comparison of treatment with a control condition
11. Selective attrition	Tendency of clients who drop out of therapy to improve less than other clients	Intent-to-treat analyses. Measurements of pretreatment differences between treatment completers versus dropouts
12. Compliance bias	Tendency for client adherence to treatment recommendations to be confounded with variables that predict improvement	Measures of treatment compliance (e.g., completion of homework assignments). Examination of compliance in the control condition
13. Selective attention to client outcomes	Tendency of individuals to unwittingly "cherry-pick" the outcome variables on which clients are improving	Blinding of observers. Explicit a priori predictions concerning client outcomes
14. Selective memory for client outcomes	Tendency of individuals to preferentially recall indications of improvement as opposed to those of no improvement or worsening	Blinding of observers
15. Selective interpretation of client outcomes	Tendency of individuals to interpret ambiguous changes in signs or symptoms as indications of improvement	Blinding of observers
Category 2 CSTEs: Misinterpretations of actual client change stemming from extratherapeutic factors		All Category 2 CSTEs: Randomization to treatment conditions
16. Spontaneous remission	Tendency of some psychological conditions to improve on their own	(none additional)
17. History	Widely shared events transpiring outside of treatment	Repeated measurements throughout treatment. Tracking of life events outside of treatment
18. Cyclical nature of some disorders	Tendency of some psychological conditions to go up and down	Long-term follow-ups
19. Self-limiting nature of disorder episodes	Tendency of individuals with psychological conditions to improve once episodes have run their natural course	Long-term follow-ups

(continued)

Table 1. (continued)

CSTE	Description	Research safeguards
20. Regression to the mean	Tendency of extreme scores to become less extreme on retesting	Use of pre- and posttest measures with high reliability. Minimizing use of extreme group designs. Estimating and controlling for measurement error
21. Maturation	Improvement arising from naturally occurring psychological growth	(none additional)
22. Multiple treatment interference	Tendency of individuals who obtain one treatment to obtain others at the same time	Use of adjunctive treatments as covariates in analyses
23. Initial misdiagnosis	Errors in the diagnosis of a condition, resulting in erroneous inferences of improvement	Use of well-validated diagnostic measures and well-trained diagnosticians. Ruling out medical causes of psychological signs and symptoms
Category 3 CSTE: Misinterpretations of actual client change stemming from nonspecific treatment factors		All Category 3 CSTE: Common factor control groups
24. Placebo effects	Improvement stemming from the expectation of improvement	Inclusion of measures of proposed mediators. Measurement of expectancies
25. Novelty effects	Improvement owing to enthusiasm regarding the prospects of receiving a new intervention	Inclusion of measures of proposed mediators. Measurement of expectancies
26. Effort justification	Clients who invest substantial time, energy, effort, and resources in an intervention often feel a psychological need to justify this commitment	Inclusion of measures of proposed mediators

Note: CSTEs in each category have one safeguard in common and then, usually, additional specific safeguards.

change stemming from extratherapeutic factors to the active treatment per se. These factors include life events that occur outside of treatment and changes in the client's psychological condition that are causally independent of treatment. In the case of Category 2 CSTEs, clients are improving, but their improvement bears no relation to either the specific or nonspecific effects of the treatment. Instead, the intervention is incidental to client change.

Category 3 CSTEs can lead individuals, again usually both clinicians and clients, to misattribute actual client change stemming from nonspecific effects of the treatment (e.g., provision of hope) to the specific effects of this treatment (see Wampold, 2001). In the case of Category 3 CSTEs, clients are improving, as they are in Category 2 CSTEs. In Category 3 CSTEs, however, this change is a consequence of common factors shared with most or all effective psychological treatments; little or none of the improvement is attributable to the specific treatment. Category 3 CSTEs are readily overlooked because they are highly correlated with the provision of the treatment. As a consequence of these CSTEs, clinicians and researchers may conclude that their hypothesized mechanisms of therapeutic effectiveness are corroborated when they are not, as these mechanisms (e.g., placebo effects) are shared by most if not all effective treatments.

Whether one regards Category 3 CSTEs as artifacts or as active agents of therapeutic change hinges largely on one's hypotheses regarding the mechanisms of improvement. If one believes that a given psychotherapy works because of specific processes that are not shared with other treatments, Category 3 CSTEs are best regarded as artifacts that can predispose to spurious inferences regarding the causes of change. In contrast, if one believes that a given psychotherapy works because of common factors that are shared with most or all effective interventions (e.g., Frank & Frank, 1961; Wampold, 2001), then the sources of change comprising Category 3 CSTEs are best regarded as valid causes of improvement in their own right. Indeed, the long-standing interest in psychotherapy integration largely reflects a desire to identify cross-cutting mechanisms that operate across many treatments (Goldfried, 2010). Hence, we caution readers against regarding Category 3 CSTEs as extraneous influences to be automatically minimized or eliminated in research, as from the standpoint of scholars who argue for the primacy of common factors in psychotherapy, such influences play a pivotal role in treatment effectiveness (e.g., Messer & Wampold, 2002).

A fourth category of inferential errors that we do not explicitly address comprises erroneous inferences regarding the mechanisms of change in a given psychotherapy. As a consequence of this class of errors, researchers and

therapists may conclude that a treatment is operating via specific mechanism X when it is actually operating via specific mechanism Y (see Kazdin, 2007). In such cases, the clinical improvements are due to specific mechanisms of the treatment but not to the specific mechanisms posited by the treatment's proponents. For example, scholars continue to debate whether cognitive-behavioral therapy works by modifying cognitions, as posited by most of its proponents (Hofmann, 2008), or by alternative mechanisms, such as increases in reinforcing activities or extinction of maladaptive thoughts and emotions (Jacobson & Christensen, 1996; Longmore & Worrell, 2007). Because the inferential errors in this fourth class involve an (a) erroneous inference regarding the specific cause(s) of treatment effectiveness rather than (b) an erroneous inference of treatment effectiveness, we do not categorize them as CSTE. In this respect, these errors differ from Category 3 CSTE, which involve the error of attributing specific effectiveness to a treatment that does not contain specific active ingredients.

Category 1 CSTE: Erroneous perceptions of client change in its actual absence

1. Illusory placebo effects. Illusory placebo effects arise when expectations of improvement lead clients to believe that an attribute or condition improves in the absence of genuine changes on specified outcome measures (Wechsler et al., 2011). Illusory placebo effects differ from placebo effects in that the former do not involve genuine change (hence, individuals harbor the illusion that they have improved when they have not), whereas the latter do.

In a clever study (Greenwald, Spangenberg, Pratkanis, & Eskenazi, 1991), experimenters switched audiotapes containing subliminal messages so that people who thought they listened to audiotapes designed to enhance memory actually listened to audiotapes designed to enhance self-esteem, and vice versa. Participants came away believing that their memory or self-esteem, as the case may be, had improved in response to the tape they believed they had heard rather than in response to the tape they had actually heard. In fact, on objective tests of memory and self-esteem, all of the tapes were ineffective. The illusory placebo effect demonstrates that expectations and implicit theories can lead people to perceive, or at least report, imaginary changes in their behaviors, thoughts, and feelings (see also Nisbett & Wilson, 1977).

2. Palliative benefits. Psychotherapy sometimes makes clients feel better about their difficulties but exerts little or no effect on these difficulties (Beyerstein, 1997). Echoing this point, Albert Ellis (2003) underscored the importance of distinguishing "feeling better" from "getting better" in psychotherapy. For example, an antisocial client may

enter therapy distressed about his repeated marital infidelity and leave therapy less distressed but with an unaltered risk for future infidelity. As Alpert (2012) observed, "Therapy sessions can work like spa appointments: They can be relaxing but don't necessarily help solve problems" (p. SR5).

One could justifiably contend that palliative changes can themselves be therapeutic in some instances, especially if distress regarding one's behaviors is a treatment target. Yet especially for clients whose behaviors routinely engender interpersonal distress for other individuals, such as those with narcissistic or antisocial personality disorders (American Psychiatric Association, 2013), the problem behaviors themselves are often the foci of the intervention. In these cases, alleviating client distress may actually be countertherapeutic. For example, some authors have argued that psychological treatment often makes psychopaths worse (Rice, Harris, & Cormier, 1992), although the research support for this contention is admittedly equivocal (D'Silva, Duggan, & McCarthy, 2004).

3. Confusing insight with improvement. Some clients may achieve greater insight into their difficulties over the course of therapy. Although such insight may not be linked to improvements in objective treatment outcomes, clients may believe that they have achieved progress merely because they can now conceptualize and verbalize their problems in greater richness and detail. In this example, insight is unrelated to improvement and thus constitutes a CSTE. If, however, the acquisition of insight per se were a therapeutic goal, then acquiring insight (even in the absence of change in signs and symptoms) would not constitute a CSTE.

There are two separable issues here, both of which bear on the veracity of insight as opposed to its clinical utility. First, the insights obtained in psychotherapy may sometimes be illusory, reflecting subjectively compelling but erroneous causal stories (Taleb, 2007). To the extent that humans are "meaning-making" beings (Bruner, 1990), insight may at times prove helpful in constructing a framework within which to better comprehend themselves and others. Indeed, some specious insights acquired in treatment may improve clients' mood or behavior, at least in the short term, by affording them a sense of understanding and control over their problems (see Jopling, 2001, for a discussion of "placebo insights" in treatment), but others may be therapeutically inert or harmful (Jopling, 2008).

Second, even if the insights accrued in therapy are veridical, they may not guarantee or even facilitate improvement. For example, a client with a specific phobia of dogs may come to recognize that his fears originated with a frightening dog attack and that he is now

negatively reinforcing these fears by avoiding dogs. Yet if he is unwilling to confront his fears during therapy by engaging in systematic in vivo exposure to dogs, his symptoms are unlikely to abate (Wachtel, 1987). Nor is insight always necessary for improvement (Bloom, 1994). In one study of psychoanalytic treatment, half of 42 patients were rated as better adjusted at the conclusion of therapy although few were judged to exhibit increased insight into their “core conflicts” (Bachrach, Galatzer-Levy, Skolnikoff, & Waldron, 1991).

4. Retrospective rewriting of pretreatment functioning. In some cases, clients may persuade themselves that they have improved by misremembering their initial level of functioning as worse than it was (Ross, 1989). Such biased memories may stem from clients’ implicit expectations of change during therapy. In one study, researchers randomly assigned university students to either a study skills course designed to improve their grades or to a no-intervention control condition and measured their study skills and grades before and after the intervention. The study skills class was apparently useless, as it failed to improve students’ grades. Yet students in the experimental condition perceived the intervention as effective, because they misremembered their initial study skills as worse than they were (Conway & Ross, 1984). Similarly, evidence suggests that at least some of the change commonly attributed to “posttraumatic growth”—psychological improvement following trauma—may actually be due to derogation of individuals’ pretrauma selves (Frazier et al., 2009; McFarland & Alvaro, 2000). Retrospective rewriting of pretreatment functioning may sometimes also occur during psychotherapy, especially when clients harbor strong expectations of improvement.

Such retrospective rewriting may transpire even when individuals are asked to evaluate their long-standing personality traits. In an elegant series of studies, A. E. Wilson and Ross (2001) found that individuals frequently described their current selves more favorably than their past selves, largely because they derogated their past selves. This tendency was especially pronounced when participants cared about the traits being judged. These results dovetail with longitudinal data demonstrating that the correlations between actual and perceived change in personality traits are only modest (Robins, Nofle, Trzesniewski, & Roberts, 2005). A study of 290 undergraduates tracked across 4 years of college found that participants retrospectively overestimated the extent to which they had become more extraverted over time, perhaps consistent with the cultural narrative that students become more outgoing and socially adept in college (Robins, Fraley, Roberts, & Trzesniewski, 2001). Such findings suggest that retrospective self-evaluations of

change in long-term therapy may be suspect, especially if they reflect implicit beliefs regarding the direction and nature of change (Ross, 1989).

5. Response-shift bias. A related phenomenon, *response-shift bias*, occurs when an intervention leads individuals to change “their evaluation standard with regard to the dimension measured” (G. S. Howard, 1980, p. 93; see also Bray, Maxwell, & Howard, 1984; G. S. Howard & Dailey, 1979). This shift, which is of particular concern for researchers or clinicians using self-report measures, can occur when an intervention leads clients to reconceptualize their initial levels of a specific psychological trait. In contrast to retrospective rewriting of pretreatment functioning, which reflects a memorial change, this CSTE reflects an alteration in one’s “implicit scale” for measuring a trait (McLeod, 2001). Response-shift bias can cause individuals to either underestimate or overestimate the effects of a psychological intervention, depending on the direction of the shift.

For example, an excessively self-critical spouse may enter couples therapy concerned that she is to blame for problems in her marriage; on self-report and interview measures, she initially rates herself as narcissistic and anger-prone. During treatment, she may come to realize that her verbally abusive and overbearing husband is primarily responsible for their marital conflicts and that her levels of self-centeredness and resentment are no higher than the average person would experience in a similarly trying situation. Even though her levels of these two problematic traits have not changed over the course of treatment, her trait scores on standardized measures may decline from pretest to posttest, leading the therapist (and often the client herself) to conclude erroneously that the treatment has lowered her self-centeredness and hostility. In a sense, the treatment has exerted an impact—on the client’s conceptualization of her traits but not on these traits themselves.¹

6. Reduction in cognitive biases. Successful treatment for depression and similar conditions may attenuate certain cognitive biases, such as those tied to self-criticism and perception of one’s level of impairment (Whisman, Miller, Norman, & Keitner, 1991). Although a reduction in such distortions is often a legitimate treatment target per se, it may engender the spurious appearance of improvement on other measures. For example, depression is often marked by overreporting of the features of associated psychopathology. As a consequence, an intervention that diminishes the intensity of the cognitive biases often associated with depression (e.g., magnification of one’s weaknesses) may lead to decreases in the reported severity of co-occurring problems (e.g., social adjustment), even when these problems have remained

objectively unchanged (Morgado, Smith, LeCrubier, & Widlocher, 1991).

7. Demand characteristics. Demand characteristics occur when clients or research participants adjust their behavior, including self-reported behavior, in accord with what they believe to be the therapists' or investigators' hypotheses (Orne, 1962). The treatment rationale provided by clinicians can convey potent demand characteristics to patients regarding treatment and thereby shape their attributions, expectations, emotions, and actions (Addis & Carpenter, 2000; McReynolds & Tori, 1972). In one study, participants informed that thoughts precede affect in response to images (i.e., a cognitive therapy rationale) were more likely to report thoughts first compared with participants informed that affect precedes thoughts. Differences between the two rationales were especially apparent in response to highly arousing images (Kanter, Kohlenberg, & Loftus, 2004) and were maintained at a 1-week follow-up (Busch, Kanter, Sedivy, & Leonard, 2007).

Moreover, clients are often motivated to tell their therapists what they believe their therapists want to hear; they may also be motivated to persuade themselves that they have improved. Hathaway (1948) referred to the "hello-goodbye" effect as clients' propensity to present themselves as worse than they actually were at the outset of treatment and better than they actually are at the conclusion of treatment. As a consequence of this phenomenon, therapists and other observers may conclude that client improvement occurred in its absence.

Similarly, hypnosis researchers have identified a "holdback effect" when participants are tested sequentially in nonhypnosis and hypnosis conditions. One of the implicit demands of hypnosis is to behave as a "good" hypnotic subject should, or at least as this role is understood by the participant (Orne, 1962). The holdback effect can arise when participants are not hypnotized during an initial baseline trial but know they will be hypnotized in the following trial. In such cases, they may deliberately "hold back" from fully responding when they are not hypnotized to demonstrate gains on the later hypnosis trial, thereby presenting themselves as good hypnotic subjects (Braffman & Kirsch, 1999; Zamansky, Scharf, & Brightbill, 1964).

8. The therapist's office error. What we term the *therapist's office error* is the mistake of confusing clients' in-session behavioral presentation with out-of-session improvement. Clients' behavior within the cloistered confines of the therapist's office may not reflect their behavior or functioning outside of treatment (Holmes, 1971; Magaret, 1950). This error may sometimes lead clinicians to underestimate treatment effectiveness, as when

adequately functioning clients use psychotherapy sessions as opportunities to express their pent-up negative emotions (see Nichols & Efran, 1985).

In other cases, however, the therapist's office error may contribute to overestimates of treatment effectiveness. For example, clients with social anxiety disorder (social phobia) involving apprehension of interpersonal rejection who are initially anxious in treatment may grow more comfortable with the therapist over time, leaving the therapist (and perhaps clients themselves) with the misleading impression that they are experiencing improvement in social anxiety symptoms. Yet these clients may merely be exhibiting stimulus discrimination, learning to respond less anxiously to the psychotherapist or others who provide them with unconditional acceptance but not to the very people they find most interpersonally threatening. Indeed, studies of behavior therapy for anxiety disorders sometimes point to a stimulus generalization gradient from the therapist's office to the outside world, reflecting marked improvements in the former setting followed by decrements upon treatment termination (Gruber, 1971; see Lynch, Chapman, Rosenthal, Kuo, & Linehan, 2006, for a discussion of real-world generalization strategies in dialectical behavior therapy). These findings underscore the need to ensure that the client's anxiety-provoking behaviors are addressed in real-world settings during treatment.

The therapist's office error may pose a particular challenge for psychoanalytic therapies, which rely heavily on the therapist-client transference as the engine of change. In many respects, one can conceptualize transferences as reflecting interpersonal expectancies (Westen, 1998). Accordingly, if clients do not generalize their positive transference reactions toward the therapist to others, their long-term improvements may be limited (Holmes, 1971).

9. Retest artifacts. The *retest artifact* (Loranger, Lenzenweger, Gartner, & Susman, 1991) is the tendency of scores on psychopathology indices to decline spuriously upon their second administration. This artifact may be especially likely with measures characterized by a skip-out structure, such as many structured and semistructured interviews. Clients may realize that if they say "no" to many questions, they will have a much briefer and less emotionally distressing experience than if they say "yes" to them, generating the false appearance of improvement. In other cases, clients may deny more symptoms during the second assessment if they learn that the questions concern sensitive behaviors, like drug use or antisocial activities. Indeed, evidence suggests that this artifact may be especially pronounced for measures of socially undesirable characteristics (Jorm, Duncan-Jones, & Scott, 1989). Although the test-retest artifact has not received the research attention it merits, data suggest

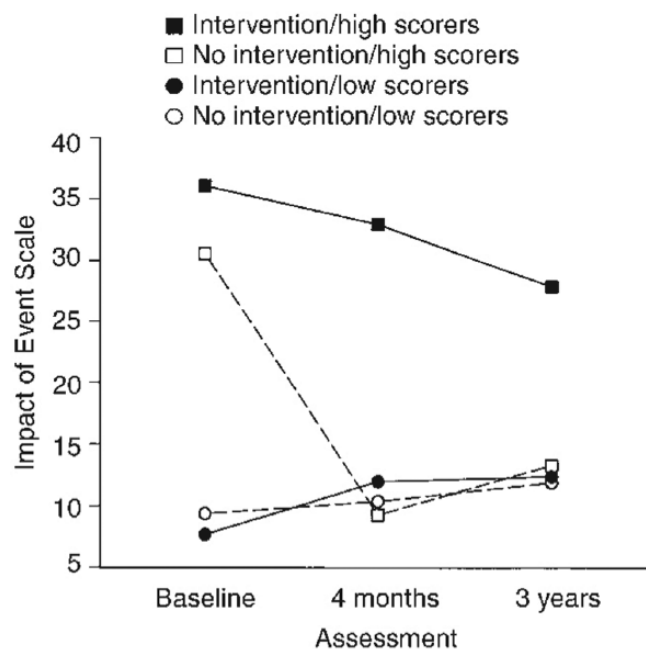


Fig. 1. The effects of critical incident stress debriefing on posttraumatic stress symptoms among traffic accident victims. Note the striking difference in trajectories between high scorers who did and did not receive the intervention. Both groups improved, but the group that received the intervention would have improved more had they received no intervention at all. From Mayou et al. (2000). Reprinted with permission.

that it may be more of a threat to the validity of short-term than long-term assessments of personality disorder features (Lenzenweger, 1999; Samuel et al., 2011).

10. Unknowable outcomes in the control condition.

A largely unappreciated reason for erroneous inferences of therapeutic effectiveness is the absence of information regarding the “hypothetical counterfactual” (Dawes, 1994): our inability to know what would have occurred had we not intervened. Because clinicians in routine practice settings are necessarily unaware of how their clients would have fared in a control condition, they cannot gauge the extent to which the improvement they observed might have occurred in the absence of treatment or in the presence of an alternative treatment. Clients are certainly subject to the same epistemic limitation.

An illustrative example derives from research on critical incident stress debriefing (CISD), which is widely used to decrease the risk of posttraumatic stress symptoms among trauma-exposed victims. Controlled research demonstrates that CISD is ineffective and perhaps iatrogenic (Litz, Gray, Bryant, & Adler, 2002; McNally, Bryant, & Ehlers, 2003). Yet many people who have undergone CISD are convinced that it was effective (Carlier, Voerman, & Gersons, 2000). A study by Mayou, Ehlers, and Hobbs (2000) offers intriguing insights into this paradox. These

investigators evaluated the 3-year outcome of 61 patients who had experienced traffic accidents; some had been randomly assigned to receive CISD and others to receive no intervention. Among other measures, participants completed the Impact of Events Scale (IES; M. J. Horowitz, Wilner, & Alvarez, 1979), an index measure of posttraumatic stress symptoms. As is evident from Figure 1, high-scoring IES participants who received CISD improved between the pretreatment baseline and the 3-year follow-up. Yet remarkably, high-scoring IES participants who received no intervention at all improved even more. These findings suggest that CISD can impede natural healing processes (McNally et al., 2003). They also help us to understand why so many people are persuaded that CISD is efficacious even though it is not. Specifically, trauma-exposed individuals who receive CISD do improve, but not because of the treatment. To the contrary, they probably would have improved even more had they received no treatment at all.

11. Selective attrition. This CSTE differs from others we have described in that it operates not at the level of individual clients but at the level of all clients in a clinician’s caseload. Selective attrition refers to the fact that clients who drop out of therapy are not a random subsample of all clients. Research demonstrates that clients who are not improving are especially likely to leave psychotherapy (Garfield, 1994; Tehrani, Krussel, Borg, & Munk-Jørgensen, 1996; see also Swift & Greenberg, 2012). As a result, therapists may conclude erroneously that their treatments are effective merely because their remaining clients are those that have improved. One problem that has long bedeviled the evaluation of Alcoholics Anonymous and similar 12-step interventions for substance disorders is the high level of client dropout from this intervention, often approaching 40% following 1 year (Kelly & Moos, 2003). The clients who remain in these treatments after several years are generally faring better than when they began, but they are unrepresentative of the clients who initially enrolled. The clients who dropped out may not have been helped or may have even been harmed by the intervention.

12. Compliance bias. A cognate problem of selection bias can arise even among clients who remain in treatment. Compliance bias occurs when differences among clients in their adherence to treatment recommendations are confounded with variables that predict outcome, such as motivation to improve or conscientiousness (Grodstein, Clarkson, & Manson, 2003; Petitti, 1994). One well-known case of such bias comes from the 1970s Coronary Drug Project, which examined the effects of clofibrate, a cholesterol-lowering medication, on heart disease (Coronary Drug Project Research Group,

1975). When the investigators detected no significant effect of clofibrate versus placebo on cardiovascular outcomes, they conducted internal analyses of regular versus irregular clofibrate users. When they did, they found that only 15% of regular clofibrate users (those who had taken 80% or more of their pills) had died of heart disease compared with 25% of irregular users, seeming to suggest a positive effect of the medication. Yet when the researchers compared regular versus irregular users of the placebo, the results were virtually identical (Dawes, 2001; Taubes, 2007). Presumably, a third variable, such as health consciousness, accounted for both (a) more diligent adherence to physicians' recommendations and (b) better cardiovascular outcomes.

Research on cognitive-behavior therapy reveals that clients who comply with extrasession homework assignments display better treatment outcomes than those who do not (Mausbach, Moore, Roesch, Cardenas, & Patterson, 2010). Similarly, evidence suggests that clients who practice meditation regularly in studies of compassion-based meditation training exhibit better outcomes than clients who do not (Pace et al., 2009). Because of compliance bias, unwary psychotherapists may notice that some of their clients comply with their prescribed interventions more than do others, find that the former clients display superior treatment outcomes, and conclude that these interventions were effective. Yet individual differences in client treatment adherence may merely be a proxy for another variable, such as treatment motivation or emotional resilience, which in turn is linked to enhanced psychological health.² Moderator analyses, which examine whether interventions are especially beneficial for certain clients (Kazdin, 2007; Kraemer, Wilson, Fairburn, & Agras, 2002), may be helpful in this regard, as levels of compliance can be treated as continuous moderators of outcome.

13. Selective attention to client outcomes. Confirmation bias (Nickerson, 1998), illusory correlation (Chapman & Chapman, 1967), and allied cognitive errors may lead clinicians to attend selectively to certain outcome variables while ignoring or minimizing others. Specifically, psychotherapists may unwittingly “cherry-pick” the outcome variables on which clients are improving. For example, because of *diagnostic overshadowing* (Garb, 1998), therapists may focus unduly on apparent improvement in dramatic client signs and symptoms, such as psychotic features or aggressive behaviors, while neglecting lack of improvement or deterioration in less overt signs and symptoms, such as depressed mood, anxious ruminations, or anger (Zimmerman & Mattia, 1999). As a consequence, they may conclude that clients have improved when they have exhibited no change across multiple clinically important domains. Clients may fall victim to

the same error: They may engage in “selective symptom monitoring” (Pennebaker & Skelton, 1981), focusing on symptoms that they expect to change while neglecting or underattending to others.

14. Selective memory for client outcomes. The past several decades of psychological research leave scant doubt that memory is fallible (Loftus, 1993; Lynn & Nash, 1994) and that most of us preferentially recall information consistent with our hunches and desires (Walker, Skowronski, & Thompson, 2003). As a consequence, clinicians may be more likely to recall positive than negative client signs and symptoms during and after treatment, potentially resulting in overestimates of treatment effectiveness.³

15. Selective interpretation of client outcomes. Confirmation bias and similar cognitive errors may predispose to selective interpretation of the clients' difficulties during and after treatment. The more ambiguous the outcome variables rated by clinicians, the larger the potential for biases in their ratings (Markin & Kivlighan, 2007; Westen & Weinberger, 2005). Hence, clinicians who are motivated to perceive improvement in their clients may interpret ambiguous symptoms (e.g., increased anger toward a spouse in marital therapy, heightened emotional processing of painful childhood memories) as evidence of treatment success.

Category 2 CSTEs: Misinterpretations of actual client change stemming from extratherapeutic factors

16. Spontaneous remission. Spontaneous remission refers to the tendency for disorders to resolve on their own (Beyerstein, 1997). Early reports by Eysenck (1952) of spontaneous remission rates of 70% or more among psychiatric patients were almost surely overestimates (Rachman, 1973). Nevertheless, later data point to nontrivial rates of spontaneous remission in outpatient samples (Chadwell & Howell, 1979; Lambert, 1976). For example, Posternak and Zimmerman (2000) reported a spontaneous remission rate of 52% among patients with major depressive disorder. The rates of spontaneous remission among children and adolescents with psychopathology, including behavioral problems, approach or exceed 40% (Harrington, Whittaker, Shoebridge, & Campbell, 1998; Jacobson & Christensen, 1996; McCullough, 2000).

The longer people remain in therapy, the greater the opportunity for extratherapeutic factors, including natural healing processes, coping, social support, and positive experiences in everyday life, to contribute to observed or perceived improvement (Jacobson & Christensen, 1996). Moreover, when frequent spontaneous remissions happen to coincide with the administration of specific

interventions, client and clinician alike may fall prey to illusory causation, coming to believe that the interventions are producing the spontaneous remissions (Blanco, Barberia, & Matute, 2014).

Spontaneous remission may be partly accounted for by what Alexander and French (1946) termed the “corrective emotional experience,” a positive affective occurrence that ameliorates the detrimental impact of early negative life events (Bridges, 2006). Although Alexander and French emphasized the role of corrective emotional experiences in psychotherapy, such events (e.g., finding a loving partner in the aftermath of an abusive relationship) surely occur in everyday life as well. In the words of psychoanalyst Karen Horney (1945), “life itself still remains a very effective psychotherapist” (p. 240).

17. History. A related extratherapeutic factor that can contribute to the erroneous inference of a therapeutic effect is what Campbell and Stanley (1963) termed *history*: widely shared events transpiring outside of the treatment setting. A client who is experiencing severe life stressors due to a poor economy or recent natural disaster may improve when the impact of these events on his (and his friends’ and loved ones’) financial and personal life has dissipated. The clinician may erroneously attribute improvement during therapy to the treatment itself rather than to the salubrious changes in the client’s everyday life.

18. Cyclical nature of some disorders. Another extratherapeutic factor that can be linked to short-term improvement is the cyclical nature of many disorders (Beyerstein, 1997). In contrast to spontaneous remission, which refers to substantial amelioration in or disappearance of a condition per se, this CSTE refers to a transient shift into the benign phase of a condition characterized by a recurrent course. Like many medical conditions, such as multiple sclerosis, arthritis, and gastrointestinal problems, many psychological disorders have their “ups and downs.” In disorders that are cyclical, people often improve, periodically or over the long term, without intervention. For example, in cyclothymic and bipolar disorders, which are characterized by affective, interpersonal, and behavioral lability, an ineffective treatment implemented over a lengthy period will have ample opportunities to coincide with upticks that likely would have occurred regardless of treatment. Accordingly, clinicians may infer that therapy is responsible for improvement when positive changes are instead induced by fluctuations in the disorders’ natural course. One likely reason for the popularity of unvalidated and fringe treatments for autism spectrum disorders, such as secretin (a polypeptide hormone synthesized from the intestines of pigs) and sensory-motor integration therapy, is the fact that the corollary symptoms

of this condition (e.g., aggression, self-injurious behavior, social interaction difficulties) often wax and wane over brief time periods (Romanczyk, Arnstein, Soorya, & Gillis, 2003), leading observers to mistake short-term behavioral changes for beneficial treatment effects.

19. Self-limiting nature of disorder episodes. Like the acute exacerbations of many physical disorders, the episodes of some psychological disorders tend to be self-limiting. A treatment may appear to exert a beneficial effect on a disorder episode that has run its natural course (Beyerstein, 1997). For example, the median duration of a depressive episode is approximately 13 weeks (Solomon et al., 2010), and some untreated episodes remit or improve substantially without any intervention (Kirsch & Sapirstein, 1998). In other cases, certain disorders themselves may be short-lived. For example, short-term drug-induced psychiatric conditions, such as amphetamine intoxication or alcohol withdrawal delirium (American Psychiatric Association, 2013), wane in intensity once the active physiological effects of the substance (or the withdrawal effects of the substance) have subsided.

20. Regression to the mean. It is a statistical fact of life that extreme scores tend to become less extreme upon retesting, a phenomenon known as *regression toward the mean* (Kruger, Savitsky, & Gilovich, 1999). By mathematical necessity, regression to the mean will occur whenever the correlation between pretest and posttest scores is less than unity (Salsburg, 2001); such regression will be especially pronounced when measures are of low reliability. If a patient presents to therapy as severely depressed, chances are reasonably high that he or she will be less depressed (or at least report lower levels of depression on standardized outcome measures) in a few weeks, even in the absence of treatment.

Regression to the mean is an especially thorny problem in evaluating the effectiveness of psychotherapy in real-world settings, because most patients enter treatment when their symptoms are most extreme and hence when regression effects are maximized (Gilovich, 1991). Similarly, antisocial children and adolescents may not be referred to treatment until their behaviors become unbearable to teachers or parents (Costello & Janiszewski, 1990). Some authors have conjectured that most of the variance commonly attributed to placebo effects in controlled trials of medication is actually due to regression effects (McDonald, Mazzuca, & McCabe, 1983). Moreover, regression effects may sometimes be misinterpreted as spontaneous remission (Campbell & Kenny, 1999).

Rendering this CSTE especially problematic are findings that humans are prone to nonregressive predictions (Dawes, 1986; Nisbett & Ross, 1980). That is, we do not sufficiently compensate for regression to the mean when

predicting behavior from Time A to Time B. Perhaps because of the representativeness heuristic, a mental shortcut characterized by the assumption that “like goes with like” (Tversky & Kahneman, 1974), we expect behavior at Time A and B to be similar. Hence, whenever we detect a difference between Time A and Time B behavior, we tend to attribute this change to spurious extraneous factors, such as the effects of treatment, rather than to statistical regression (an error known as the *regression fallacy*; Kahneman, 1965). As Campbell and Kenny (1999) commented, “it seems likely that regression toward the mean leads people to believe in the efficacy of the scientifically unjustified regimens. . . . Many a quack has made a good living from regression toward the mean” (p. 48).

21. Maturation. A source of erroneous inferences of therapeutic efficacy, especially among children and adolescents, is maturation: improvement owing to naturally occurring psychological growth (Cook & Campbell, 1979). For example, children and young adolescents with high levels of what appear to be certain pre-psychopathic features, such as poor impulse control, low frustration tolerance, and defiance, may improve on their own because levels of these characteristics often diminish with the passage of time, especially when they are early appearing (Edens, Skeem, Cruise, & Cauffman, 2001). Such maturation can mislead clinicians into concluding that their treatment was responsible for declines in the levels of these and other externalizing problems. Psychological growth may be a source of mistaken therapeutic conclusions even among adult clients. For example, some patients with borderline personality disorder may improve over long stretches of time without treatment (Shea et al., 2009).

22. Multiple treatment interference. When clients seek out a treatment, they often obtain other interventions simultaneously (Kendall, Butcher, & Holmbeck, 1999), a confound known as *multiple treatment interference* or *co-intervention bias*. Some of these adjunctive interventions may be formal treatments, such as antidepressants or marital therapy. Others may be informal “treatments,” such as exercise, which has generally been found in controlled studies to be effective for alleviating depression (Fremont & Craighead, 1987; Penedo & Dahn, 2005), or confiding in trusted friends or religious figures. Multiple treatment interference renders it difficult or impossible to attribute client change conclusively to the active ingredients of the intervention of choice.

23. Initial misdiagnosis. Even the best trained diagnosticians are fallible (Beyerstein, 1997; Garb, 1998; Groopman, 2007). For example, relatively normal individuals

undergoing temporary life stressors are at times mistakenly diagnosed as psychopathological; when they are later examined, they have improved but not necessarily because of the treatment. The same may hold for clients with acute medical disorders that are misdiagnosed as psychiatric conditions. For example, acute intermittent porphyria has been called “a great imitator” (Morrison, 1997, p. 155) and is occasionally mistaken for bipolar disorder and other cyclical emotional conditions. If this medical condition resolves on its own, which it sometimes does (Loftus & Arnold, 1991), an unwary clinician may mistakenly conclude that a treatment targeted for a manic episode was beneficial.

Category 3 CSTEs: Misinterpretations of actual client change stemming from nonspecific treatment factors

24. Placebo effects. The omnipresent placebo effect has been defined in multiple ways, but it is traditionally regarded as improvement resulting from the mere expectation of improvement (Beecher, 1955; S. Horowitz, 2012; Steer & Ritschel, 2010). By instilling hope and the conviction that one can rise above life’s challenges, virtually any credible treatment can be at least somewhat helpful for combating demoralization (Frank & Frank, 1961), which is a central component of many psychological disorders (Tellegen et al., 2003). Admittedly, importing the placebo concept into the domain of psychotherapy is fraught with complexities given that at least some of the efficacy of psychological treatment probably derives from expectancies of improvement (Kirsch, 2005; Lambert, 2005). Nevertheless, because such expectancies presumably cut across most or all effective psychotherapies, they can lead clinicians and researchers to conclude that the specific ingredients of a treatment are efficacious when they are inert.

In the case of medication, some research suggests that up to 80% of the effects of antidepressants on clinical depression, especially when it mild or moderate, may be attributable to placebo effects (Kirsch, 2005; Kirsch & Sapirstein, 1998; but see Coyne, 2012; Klein, 1998, for different views). Placebos generally exert their most potent effects on subjective reports, such as depression, pain, and nausea, rather than on largely objective indices, such as assays of cancer, heart disease, or other organic illnesses (Hróbjartsson & Gotzsche, 2001).

Placebo effects appear to play an important role in the efficacy of psychotherapy, too. Estimates of placebo effects in psychotherapy, typically obtained by comparing treatment outcomes from attention-placebo control groups with those of wait-list control groups, are on the order of $d = 0.40$, or about half of the typical effect size yielded by active therapies (Grissom, 1996; Lambert, 2005; Lambert & Ogles, 2004). Moreover, meta-analyses

indicate that the estimated efficacy of psychotherapy is considerably smaller when it is compared with an attention-placebo control group than with a wait-list control group (Baskin, Tierney, Minami, & Wampold, 2003; Bowers & Clum, 1988), suggesting that some of the potency of psychological treatment derives from the non-specific effects of expectancies.

Still, ascertaining the precise magnitude of placebo effects in psychotherapy is difficult and arguably impossible given the absence of a perfect psychological treatment analogue to a pill placebo (Kirsch, 2005). Moreover, even control conditions designed to be active placebos may not control fully for the effects of expectancies, as these placebos are often less plausible than the active interventions against which they are pitted (Boot, Simons, Stothart, & Stutts, 2013).

Placebo effects should not be confused with other nonspecific effects of treatment (Kienle & Kiene, 1997; cf. Novella, 2010), such as those of empathy and support (Nathan, Stuart, & Dolan, 2003). The causal role of these nonspecific factors is controversial. On the one hand, the therapeutic alliance is modestly and positively associated (average $r = .22$) with therapeutic improvement (Baldwin, Wampold, & Imel, 2007; Orlinsky, Rønnestad, & Willutzki, 2004). This finding has led some scholars to contend that the therapeutic alliance is a causal agent in psychotherapeutic change. On the other hand, relatively few therapy outcome studies account for the temporal relation between the alliance and improvement, precluding relatively clear-cut inferences of causality (Kazdin, 2007). Several investigations that have incorporated assessments of therapeutic alliance and symptom change at multiple therapeutic time points suggest that a positive alliance typically follows symptom change, not vice versa (DeRubeis, Brotman, & Gibbons, 2005; DeRubeis & Feeley, 1990); but other studies have arrived at different conclusions (Horvath, Del Re, Flückiger, & Symonds, 2011; Norcross & Lambert, 2006). In light of this mixed evidence, we do not class these nonspecific factors as CSTE.

25. Novelty effects. Clients may improve, especially at the outset of treatment, because they are excited by the prospect of receiving a new intervention (Fraenkel & Wallen, 1993; Marino & Lilienfeld, 2007). Novelty effects probably overlap with placebo effects in some cases, but the former typically operate largely or exclusively during the initiation of treatment.

Psychotherapy outcome data suggest that about 15% of patients improve between the initial phone call from the clinician and the first session (K. I. Howard, Kopta, Krause, & Orlinsky, 1986). At least some of this improvement probably stems from the anticipation of receipt of a novel treatment. Moreover, for many conditions,

including major depression and eating disorders, perhaps 60% to 80% of clinical improvement occurs by the fourth session (Ilardi & Craighead, 1999; G. T. Wilson, 1999; but see Tang & DuRubeis, 1999). Much of the early change in psychotherapy may similarly reflect clients' reactions to a new intervention that offers the promise of change, although some of it may also stem from placebo or regression effects. Novelty effects may account in part for meta-analytic findings that the effect sizes for the efficacy of some psychotropic medications, including second-generation antipsychotics and antidepressants, have been highest shortly following their introduction, only to dissipate with time (Lehrer, 2010; Leucht, Arber, Engel, Kissling, & Davis, 2009), although other factors (e.g., enrollment of progressively milder patients in medication studies, publication bias) may also be at play.

26. Effort justification. Because clients often devote substantial time, energy, effort, and money to treatment, they may feel a need to justify this investment. They may do so by persuading themselves that the therapy was beneficial, a phenomenon termed *effort justification* (Cooper, 1980; Cooper & Axsom, 1982). In one study, college students with snake phobic symptoms improved equally when receiving exposure therapy and when performing strenuous physical exercises (e.g., running quickly in place). The latter "treatment" required considerable effort and presumably led to a need to rationalize this effort (Axsom & Cooper, 1985). Effort justification may be a particularly challenging interpretative problem for long-term insight-oriented therapies, especially those lasting decades, because of the enormous financial, time, and emotional investment involved.

Summary

These 26 CSTE are a helpful springboard for examining why certain inert or harmful treatments (or treatment ingredients) may appear to be effective. Our list is only a starting point, however, because CSTE almost surely comprise only one set of sources for incorrect inferences regarding treatment effectiveness. Other sources include the fact that clinicians are often extremely busy and are therefore forced to make rapid decisions in complex and information-rich environments. Moreover, as noted earlier, another source of erroneous inferences comprises incorrect hypotheses regarding the specific mechanisms of a treatment. We encourage additional research on other potential CSTE, as well as on shared processes that may underpin superficially different CSTE.

Like many rival hypotheses in psychology (Huck & Sandler, 1979), CSTE are readily overlooked because they are nonintuitive. In addition, they are less perceptually obvious than the easily observed impact of client

change and therefore are likely to recede into the causal background (Lilienfeld et al., 2008). As a consequence, some clinicians may assume erroneously that they can dispense with the research leg of evidence-based practice and replace it with informal clinical observation.

Research Methods as Safeguards Against Causes of Spurious Therapeutic Effectiveness

A key point that is not emphasized sufficiently in education in clinical psychology and allied disciplines is that systematic research designs, including both between-subject and single-subject designs, are needed to minimize CSTE as rival hypotheses for client improvement (Lilienfeld et al., 2008; T. D. Wilson, 2011). In many respects, the existence of CSTE offers the most potent *raison d'être* for evidence-based practice, although to our knowledge this crucial point has never been made explicitly (but see Lilienfeld et al., 2013; and Stewart et al., 2011, for discussions of sources of resistance toward evidence-based practice).

Specifically, without randomized controlled trials, well-controlled quasi-experimental studies, systematic single-subject designs, and other research methods as partial safeguards against CSTE, there is no way to ascertain whether client change was due to the intervention as opposed to a wealth of extraneous factors. Randomized controlled trials are not strict “gold standards,” as they do not remove all potential sources of error (Wachtel, 2010). Nevertheless, analyses of the medical literature suggest that treatment designs based on random assignment tend to yield more replicable results than do those based on quasi-experimental or naturalistic designs (Ioannidis, 2005), probably at least in part because the former help to eliminate more CSTE as rival explanations for improvement.

As a consequence of their superior control over CSTE, randomized controlled trials and rigorous single-subject designs justifiably occupy the highest rungs of evidentiary certainty in the evidence-based practice hierarchy (Ghaemi, 2009). Nevertheless, designs lower in this hierarchy, such as quasi-experimental and naturalistic methods, can also play valuable roles in research inference, as they help to protect investigators against certain CSTE (Wachtel, 2010). Moreover, such designs are often indispensable in the early phases of treatment development, as they allow researchers to collect preliminary data that can shape the development of novel interventions. In turn, these interventions, if feasibly implemented and empirically promising, can later be tested in more rigorously controlled trials.

In the next section, we sketch out how widely used methodological procedures in psychotherapy outcome research help to eliminate or minimize CSTE. Our

exposition is instructive rather than exhaustive; we focus only on the most crucial research safeguards and most crucial CSTE (again see Table 1 for these and additional methodological safeguards against CSTE).

Protecting against Category 1 CSTE

Well-validated outcome indicators. Well-validated and largely objective outcome measures help to rule out all Category 1 CSTE, because these CSTE can engender the false appearance of change in its absence. For example, well-validated indicators of depression or anxiety help to exclude—although not eliminate—illusory placebo effects and palliative effects in controlled trials of major depression and anxiety disorders. To be effective safeguards against Category 1 CSTE, well-validated outcome indicators should be sensitive not only to client symptoms but also to client impairment. Such indicators are also useful as protections against Category 1 CSTE in controlled single-subject designs. In contrast, demand characteristics can be especially difficult to rule out as sources of erroneous clinical inference. Nevertheless, outcome measures that are low in reactivity (Weiss & Weisz, 1990), such as extrasession behavioral data or unobtrusive behavioral observations, are at least partial antidotes against this CSTE. Collateral reports from outside informants (e.g., friends, significant others), which can supply “social validation” (Kazdin, 1977), can be useful in ruling out the confusion of insight with improvement, retrospective rewriting of pretreatment functioning, response shift bias, the therapist’s office error, and similar CSTE. Specifically, these reports can assist clinicians and investigators with excluding the hypothesis that client-perceived change in symptoms is (a) limited to behaviors within therapy sessions, (b) illusory, or (c) both.

Pretreatment measures. Collecting measures of client psychological status at pretreatment is especially helpful for ruling out one specific Category 1 CSTE, namely, retrospective rewriting of pretreatment functioning. Specifically, such measures can assist in excluding the hypothesis that clients are merely misremembering their initial adjustment as worse than it actually was, thereby leading to spurious inferences of improvement. If these measures do not rely exclusively on self-report ratings, they can also help to eliminate response-shift biases as explanations for apparent improvement.

Blinding of observers. Blinded observations in controlled clinical trials control partially for several additional Category 1 CSTE, especially those stemming from confirmation bias and illusory correlation (i.e., selective attention, memory, and interpretation of client outcomes). When external evaluators are fully blinded, they cannot

subtly and selectively perceive, recall, or interpret ambiguous symptom changes as a function of treatment assignment. For example, blinded observers in a randomized controlled trial of cognitive-behavior therapy versus a wait-list control for generalized anxiety disorder are less likely to differentially elicit or cherry-pick indicators of improvement (e.g., reports of less frequent worrying) in the treatment condition.

Nevertheless, these Category 1 CSTE may be difficult to eliminate entirely. Because therapy outcome studies cannot be strictly double-blinded (i.e., clients and clinicians know who is receiving treatment), confirmation bias can still affect ratings of improvement by clients and clinicians. Moreover, even the blinding of external observers in psychotherapy trials is rarely infallible, as these evaluators can often surmise treatment assignment at above-chance levels (Carroll, Rounsaville, & Nich, 1994). Assessing potential violations of blinding by asking evaluators to guess treatment conditions and using this variable as a covariate in analyses can be a helpful safeguard against selective perception, memory, and interpretation of client change. Nevertheless, such covariate analyses may underestimate treatment differences (especially when based on guesses made at the conclusion of treatment), because above-chance guessing could stem from evaluators' accurate observations of differential improvement across conditions (Carroll et al., 1994; Rickels, Lipman, Fisher, Park, & Uhlenhuth, 1970).

Intent-to-treat analyses. Intent-to-treat (ITT) analyses (Hollis & Campbell, 1999) help to rule out one key Category 1 CSTE, namely, selective attrition. By examining outcomes of all participants enrolled in clinical trials, including dropouts, ITT analyses minimize erroneous inferences of improvement stemming from the fact that clients who leave treatment prematurely are often unrepresentative of those who initially enrolled (Tehrani et al., 1996). In contrast to clients who remain in treatment, those who drop out of treatment tend to be lower functioning and more psychologically disturbed (Swift & Greenberg, 2012), although in a minority of cases they comprise clients who have improved and no longer perceive themselves as requiring treatment (Baekeland & Lundwall, 1975; Tehrani et al., 1996). As a consequence of the selection biases introduced by client dropout, ITT analyses help to avoid misestimating—and typically overestimating—treatment effects.

Protecting against Category 2 CSTE

Randomization to treatment conditions. Randomization to treatment conditions helps to address the inferential errors generated by Category 2 CSTE, which produce changes stemming from extraneous factors outside of treatment. To be clear, well-executed randomized

controlled trials do not eliminate Category 2 CSTE, which still arise in these investigations and can deceive observers in the absence of randomized controlled groups. Nevertheless, the randomization process helps to exclude Category 2 CSTE as rival explanations for therapeutic effectiveness, because these CSTE are equally likely in sizeable experimental and control groups. Given the law of large numbers, these CSTE should no longer account for between-group differences in randomized controlled trials provided that clinical trials are adequately powered (Hsu, 1989). For example, in a randomized controlled trial, spontaneous remission, history, regression to the mean, maturation, and multiple treatment interference occur frequently among individuals assigned randomly to both treatment and no-treatment (or alternative treatment) conditions. Nevertheless, proper randomization ensures that these CSTE tend to be equalized across the active treatment and comparison arms.

Repeated measurements. In both between-subject and controlled single-subject experiments, repeated measurements across the course of treatment can help to rule out history and other extratherapeutic influences as sources of improvement in therapy (Laurenceau, Hayes, & Feldman, 2007). If one observes changes in treatment at multiple time points rather than at only one time point following an extratherapeutic event (e.g., initiation of a romantic relationship), the likelihood that such events—rather than the therapeutic intervention—are contributing to improvement is minimized (such observations are also useful for ruling out novelty effects, a Category 3 CSTE). In the context of single-subject designs, multiple baseline designs—especially those in which the intervention is applied to different behaviors in a temporal sequence—can help to rule out history and other extratherapeutic factors as rival explanations for change during treatment (Engel & Schutt, 2012; Nock, Michel, & Photos, 2007). If one consistently observes change in different behaviors at different time points, the likelihood that extratherapeutic factors account for the improvement is minimized. Finally, long-term follow-up measurements can be helpful in excluding CSTE arising from the cyclical and self-limiting nature of certain disorders, as such assessments can ensure that improvements in signs and symptoms are not transient.

Minimizing and estimating measurement error.

The use of pre- and posttreatment indicators with high reliability will minimize regression to the mean, as this statistical phenomenon is most probable when measures contain substantial amounts of nonsystematic (random) measurement error. Particularly in quasi-experimental treatment studies, investigators should be circumspect in their use of extreme-groups designs (in which participants are selected on the basis of very high pretreatment

scores), as such designs are especially likely to yield high levels of regression effects. Researchers can also estimate, and control statistically for, regression effects in treatment outcome studies (see Barnett, van der Pols, & Dobson, 2005, for a discussion).

Protecting against Category 3 CSTEs

Common factor control groups. Systematic controls for common therapeutic factors help to control for Category 3 CSTEs, which involve the misattribution of client change to specific therapeutic ingredients when common factors (e.g., expectancies for improvement) are actually operative. For example, as observed earlier, attention-placebo control groups (Paul, 1966) can minimize expectancies for change as an explanation for group differences in outcome. Nevertheless, attention-placebo control groups are unlikely to eliminate entirely the threats posed by Category 3 CSTEs, because even well-constructed common factor control conditions are rarely as plausible as active treatment conditions (Baskin et al., 2003; Boot et al., 2013; O’Leary & Borkovec, 1978). Given the difficulty and perhaps impossibility of equating common factor control conditions with active treatment conditions on expectancies for change, hope, and treatment credibility, researchers and clinicians should ideally measure these factors at different points in treatment (in the case of novelty effects, at the outset of treatment). These variables can be treated as covariates in analyses, again bearing in mind that such statistical controls can underestimate treatment effects if expectancies and treatment credibility in part contribute to treatment efficacy.

Inclusion of measures of proposed mediators. The demonstration that a proposed mediator of treatment outcome accounts statistically for client improvement supports, although does not prove, the contention that this mediator is the underlying mechanism of change (see Kazdin & Nock, 2003, for conditions in which mediation offers especially compelling evidence for change mechanisms). In this regard, mediational tests can be helpful for excluding Category 3 CSTEs. Specifically, converging findings that a given psychotherapy appears to operate via a hypothesized mediator that is largely specific to that intervention (e.g., changes in maladaptive cognitions, cognitive defusion, increase in social reinforcement) minimizes the likelihood that this intervention is operating exclusively via common mechanisms shared by most or all treatments, such as placebo effects.

Summary

Methodological techniques in psychotherapy outcome research help to control for CSTEs, and certain methods

are especially suited for excluding different CSTEs. The need to minimize CSTEs using between-subject and single-subject research designs offers the most compelling rationale for evidence-based practice. Our discussion also points to important gaps in methodology for attenuating the influence of CSTEs as well as fruitful directions for future research. As is evident from our analysis, Category 3 CSTEs are especially difficult to eliminate as erroneous sources of improvement, because equating active treatment and attention-placebo control conditions on expectancies and treatment credibility is often difficult or impossible. Hence, one key direction for future psychotherapy research will be the development of placebo conditions that are closely matched to treatment conditions on credibility (see Boot et al., 2013). Moreover, because psychotherapy trials cannot be conducted in a genuinely double-blind fashion, certain Category 1 CSTEs, especially selective attention, memory, and interpretation of client outcomes, are difficult to eradicate, particularly for client and clinician reports of improvement. The development of largely objective measures that are less susceptible to these and other observer biases is therefore an important direction for future psychotherapy outcome research. In the case of all CSTE categories, researchers are well advised to heed the methodological maxim that if one cannot remove a source of error, one should attempt to measure it. For example, by systematically assessing expectancies during treatment, investigators can strive to rule out rival hypotheses concerning client improvement and thereby draw more valid inferences regarding treatment effects.

Conclusions and Future Directions

The oft-lamented gap between science and practice in clinical psychology is in large measure a clash of epistemologies (McHugh, 1994). In particular, this schism reflects deep-seated differences of opinion regarding the place of controlled research versus intuition in clinical decision making (Lilienfeld et al., 2013; Tavris, 2003). Our central thesis is that the science–practice gap and the accompanying reluctance of some psychologists—clinicians and researchers alike—to adopt evidence-based practices rarely reflect a willful disregard of evidence per se. Instead, this reluctance stems largely from an erroneous belief that the evidence supplied by informal clinical observations of client change tends to be as trustworthy as the evidence supplied by the methodological safeguards comprising the research prong of evidence-based practice (Spring, 2007). When viewed through this lens, the science–practice gap is not fundamentally a disagreement about whether evidence is important in ascertaining therapeutic effectiveness: It is a

difference of opinion about which kinds of evidence should be accorded priority in clinical decision making.

Implications for the role of intuition in clinical decision making

Clinical intuitions and informal observations play invaluable roles in psychotherapy, especially in hypothesis generation (Chambless, in press). For example, the spark that ignited Aaron Beck's seminal theorizing regarding cognitive-behavioral therapy originated from his observations of a client who seemed anxious during sessions. After Beck, who was trained psychoanalytically, suggested to her that her anxiety reflected discomfort with unconscious sexual impulses, she replied politely that she felt nervous because she was concerned she was boring him. This experience inspired Beck to explore his clients' unstated thoughts and assumptions, culminating in the development of what he initially termed *cognitive therapy* (Smith, 2009). Moreover, clinical impressions of change during treatment are sometimes accurate and should be regarded as fallible but potentially informative signposts to be corroborated by more systematic evidence. At the same time, our analysis is a reminder that clinical observations are often poorly suited to detecting and evaluating the sources of improvement in treatment. The evidence we have reviewed demonstrates that (a) throughout history, ineffective and harmful mental health treatments have routinely been perceived as effective; (b) psychotherapists frequently overestimate substantially the rates of positive outcomes in their clients (Hannan et al., 2005); and (c) many sources can contribute to the erroneous impression of therapeutic effectiveness in its absence.

One potential response to our arguments is that CSTE are less of an impediment for highly experienced psychotherapists, who can gradually learn to distinguish accurate from inaccurate inferences of treatment effectiveness. Nevertheless, research across multiple domains reveals that the conditions for the acquisition of intuitive expertise are highly constrained. Intuitive expertise tends to emerge only in "high-validity environments"—those in which feedback is relatively objective, consistent, and immediate (Dawes, 1994; Kahneman & Klein, 2009; Tracey, Wampold, Lichtenberg, & Goodyear, 2014). None of these conditions apply to typical psychotherapy, a "low-validity environment" in which feedback to clinicians is often ambiguous (e.g., detecting whether a client is less anxious than in the previous session can be challenging, and detecting whether such change is due to the intervention itself is even more so), inconsistent (e.g., a client may appear improved in one session but not in the succeeding session), and delayed (e.g., clinicians may

need to wait weeks or months before discovering whether a client improved following an intervention). Hence, the literature on expertise effects provides scant reason to expect the accuracy of intuitions concerning therapeutic effectiveness to improve with experience. Research on the relation between the amount of therapeutic experience and accuracy of clinical judgments offers few (Spengler et al., 2009) or virtually no (Garb, 1998, 2005) additional grounds for optimism.

Implications for prioritizing sources of evidence

These points bear noteworthy implications for the weighting of the three legs of evidence-based practice: research evidence, clinical expertise, and client preferences and values (Spring, 2007). Although some authors contend or imply that these three prongs should be accorded approximately equal weight in clinical decision making (e.g., American Psychological Association Task Force on Evidence-Based Practice, 2006), our analysis suggests that this ecumenical approach may be misguided. As we have seen, controlled research on treatment efficacy is better suited than unguided clinical judgment to ruling out manifold rival explanations for improvement, a finding that accords with the superior replicability of medical findings derived from randomized controlled trials compared with less rigorously controlled trials (Ioannidis, 2005). Hence, when well-replicated treatment outcome data conflict with clinical impressions of improvement, we should generally default to the former (Baker et al., 2008).

Implications for everyday clinical practice

Our arguments point to useful suggestions for everyday clinical practice as well (see Table 1). Because of Category 1 CSTE, clinicians can be led to conclude that client change has occurred when it has not. One underutilized corrective to this problem is the periodic administration of outcome measures, such as the Outcome Questionnaire-45 (Lambert, Lunnen, Umphress, Hansen, & Burlingame, 1994), throughout treatment. These measures can alert clinicians to instances in which they may be erroneously perceiving improvement in its absence or overlooking deterioration. Some Category 1 CSTE, especially those stemming from confirmation bias on the part of both clinician and client, can be minimized by collecting systematic data from outside informants. In some cases, as in the treatment of anxiety disorders, these CSTE can also be minimized by collecting psychophysiological data (e.g., autonomic responsivity to anxiety-provoking events) over the course of treatment.

Although many Category 2 and 3 CSTE are difficult to rule out in everyday clinical practice, clinicians can often enhance the accuracy of their inferences of treatment effectiveness by inquiring about these CSTE systematically. For example, by monitoring clients' use of informal adjunctive "interventions," such as exercise, herbal remedies, and confiding in valued friends, clinicians can become more cognizant of multiple treatment interference as a potential explanation for changes in their clients' clinical status.

Our analysis also reminds clinicians to be attuned to the possibility that some client characteristics may serve as moderators of certain CSTE, thereby affecting their likelihood. For example, as noted earlier, people with depression may be especially prone to symptom over-reporting and hence to the false appearance of improvement on adjunctive symptoms following treatment (Morgado et al., 1991). Similarly, because individuals with high levels of negative emotionality, especially trait anxiety, are prone to attend selectively to psychological symptoms (Suls & Howren, 2012), clinicians should be alert to the possibility that declines in negative emotionality over the course of treatment could predispose to spurious inferences of declines in other psychological symptoms. In addition, although efforts to identify a "placebo-prone personality" have met with mixed success, some evidence raises the possibility that optimists are more likely than pessimists to respond to positive expectancies (Geers, Helfer, Koskab, Weiland, & Landry, 2005) and hence may be especially prone to engendering certain CSTE, especially placebo and novelty effects.

Implications for clinical psychology education and training

Our analysis implies that CSTE and the research safeguards against them that we have delineated should be emphasized in the education and training of all would-be psychologists and other mental health professionals, as well as in the continuing education of current mental health professionals. Although we are unaware of survey data on how often CSTE are discussed in graduate courses in the mental health professions, there is reason to believe that such coverage is often minimal. Most standard psychotherapy handbooks (e.g., Corsini & Wedding, 2010; Koocher, Norcross, & Hill, 2005; Meyer & Deitsch, 1996) accord scant attention to the overarching problem of CSTE or to specific CSTE themselves, such as placebo effects, spontaneous remission, and regression to the mean. Moreover, to our knowledge, no continuing education course approved by the American Psychological Association has ever focused on CSTE.

In our view, the inferential problems posed by CSTE, and, equally important, the ways in which research safeguards compensate for them, should be mandatory components of training for all mental health professionals. Exposing students to the long and sordid history of failed but widely espoused treatments in psychology and psychiatry may be especially helpful as a didactic device. In addition, a thoughtful consideration of CSTE should be integrated routinely into clinical supervision and case presentations. For example, when reviewing client improvement over the course of treatment, supervisors should encourage trainees to carefully consider (a) rival explanations for such improvement other than, or in addition to, the intervention itself, (b) cognitive biases that may lead to false inferences in this regard, and (c) safeguards against drawing erroneous inferences concerning the existence and sources of improvement. Nevertheless, because it is not known whether instruction in CSTE enhances therapeutic outcomes, we call for research on this question. The absence of such evidence notwithstanding, education regarding CSTE may diminish resistance to evidence-based practice among future and current clinicians, as such knowledge provides a persuasive rationale for reliance on research designs to gauge therapeutic efficacy and effectiveness (Lilienfeld et al., 2013).

Our arguments also bear implications for training models in clinical psychology. For example, the local clinical scientist model (Stricker & Trierweiler, 1995), adopted by many or most scholar-professional (Psy.D.) programs, encourages clinicians to operate as scientists within the miniature laboratory of the clinical setting, carefully observing client behaviors in response to interventions, generating hypotheses about them, and testing these hypotheses with additional interventions. In principle, these are laudable goals. Nevertheless, CSTE raise largely unappreciated challenges for the implementation of the local clinical scientist model, because they render it difficult to draw reasonably clear-cut conclusions regarding treatment effectiveness for individual clients. Hence, although local ("idiographic") clinical science certainly has its merits, it cannot substitute for nomothetic clinical science derived from randomized controlled trials, single-subject designs, and other systematic research methods.

At the same time, idiographic clinical science is hardly a dead end, so clinicians need not despair. Although clinicians operating in the context of individual clients cannot exclude many alternative explanations for client improvement, especially Category 3 CSTE, they can nonetheless evaluate client change through the prism of CSTE. In this way, they can become more alert to alternative explanations for change. For example, as noted earlier, clinicians can monitor client change systematically across sessions, thereby permitting them

to minimize illusory placebo effects; they can solicit information from informants regarding clients' out-of-session behaviors to minimize the therapist's office error; they can attend diligently to all relevant client outcomes to avoid inadvertent cherry-picking of signs and symptoms; they can be alert to the fact that client improvements over time may reflect regression to the mean, history, and other artifacts; they can attend to potential client characteristics that may moderate the likelihood of CSTE; and so on. In this respect, they can adopt a scientific mind-set while bearing in mind that various sources of inferential error cannot be completely eliminated. Hence, the local clinical scientist model, although not an adequate substitute for scientist-practitioner or clinical science models of training (see Baker et al., 2008), is a helpful reminder that clinicians should continually operate as "detectives" who strive to identify potential rival sources of improvement and who (a) minimize these sources when they can and (b) bear them in mind as inferential constraints when they cannot.

Limitations of our analysis

Our analysis is limited in at least three respects. First, our review leaves unresolved the question of how often each of the 26 CSTE we have identified contributes to erroneous inferences in actual clinical practice. As in many domains of psychology, one must distinguish "can" from "does" in discussions of causality (McCall, 1977). The fact that a CSTE can lead to incorrect inferences of therapeutic effectiveness does not tell us how often it does so. Research examining therapists' knowledge of and understanding of CSTE, both in the abstract and in real-world practice, would be a useful starting point in addressing this question.

Second, we have focused only on inferential errors that apply to everyday clinical practice. We have not examined the many methodological decisions that can generate spurious inferences of treatment effectiveness in research studies of all kinds. For example, the file-drawer effect (Rosenthal, 1979), which is the bias against submitting negative results for publication, and outcome reporting bias (Chan & Altman, 2005), which is the propensity to cherry-pick data on dependent measures that yield positive results, can lead to overestimates of treatment efficacy. Recent data also raise the possibility of a disconcertingly high prevalence of "*p*-hacking," that is, analyzing data—or peeking repeatedly at already collected data—until alpha levels fall just below .05 (Masicampo & Lalande, 2012). More directly relevant to clinical practice, some researchers also contend that the use of wait-list control groups contributes to overestimates of psychotherapy efficacy, because clients in these groups may deteriorate as they await treatment, experience "resentful

demoralization" (Cook & Campbell, 1979) as a consequence of not receiving treatment afforded to other individuals, or both. Nevertheless, the evidence for this assertion is mixed (e.g., S. A. Elliott & Brown, 2002). Still other authors argue that "treatment as usual" conditions, which often serve as control groups in psychotherapy outcome designs, are best conceived of as "intent to fail" conditions, defined (perhaps tendentiously) as "pseudotreatments designed specifically as control groups to prove the superiority of the investigator's preferred treatment and that have no theoretical rationale or are delivered by graduate students who know they are administering treatment that is not supposed to work" (Westen & Bradley, 2005, p. 267). If these critics are correct, some standard psychotherapy outcome designs may overestimate the efficacy of beneficial treatments or generate the mistaken conclusion that inefficacious treatments are efficacious.

Third, some readers might contend that our core arguments are rendered effectively moot by the Dodo Bird verdict, named after the Dodo Bird in Lewis Carroll's *Adventures of Alice in Wonderland*, who proclaimed following a race that "Everybody has won, and all must have prizes." This verdict posits that all psychotherapies are (a) effective and (b) equivalent in their effectiveness, both overall and for all disorders (Luborsky, Singer, & Luborsky, 1975; Shedler, 2010; Wampold et al., 1997). If the Dodo Bird verdict is correct, the reasoning continues, CSTE are of little or no concern because all treatments work, and work equally well (see Stewart et al., 2011).

Nevertheless, the Dodo Bird verdict has historically referred to a rough equivalence in the effectiveness of different schools of therapy (e.g., psychodynamic, cognitive-behavioral) rather than to a precise equivalence in the efficacy of all specific treatments (e.g., Smith, Glass, & Miller, 1980). Moreover, the assertion that all therapies are of equal efficacy, either overall (a main effects hypothesis) or for all conditions (an interactional hypothesis), is difficult to sustain (Lilienfeld, 2014; but see Wampold, 2001, for a more sanguine perspective on the Dodo Bird verdict). For example, well-replicated data indicate that exposure-based therapies are more efficacious than other treatments for at least some anxiety-related disorders (e.g., obsessive-compulsive disorder) and that behavioral therapies are more efficacious than nonbehavioral therapies for child and adolescent behavioral problems (Chambless & Ollendick, 2001; Hunsley & Di Giulio, 2002). A meta-analysis by Tolin (2010) similarly revealed that behavioral and cognitive-behavior therapies are more efficacious than other therapies for anxiety and mood disorders. Further calling into question the Dodo Bird verdict are findings that at least some interventions, such as CISD, are at best ineffective and perhaps harmful (Lilienfeld, 2007; McNally et al., 2003). Even Bruce

Wampold, a prominent proponent of the Dodo Bird verdict, acknowledges that this conclusion applies only to “bona-fide” therapies, namely, those based on sound psychological principles, delivered by well-trained psychotherapists, and laid out explicitly in manuals or other publications (Wampold & DeFife, 2010).

Furthermore, our discussion of CSTE is relevant not only to “schools” of psychotherapy but also to specific therapeutic techniques, many of which transcend diverse treatment modalities. That is, many CSTE can predispose to false inferences regarding the effectiveness of specific techniques delivered within a therapy session, such as an interpretation of a client’s statement, a piece of advice imparted to a client, or a role-play exercise between client and clinician. Hence, even setting aside the contested Dodo Bird verdict, the inferential problems posed by CSTE in the psychotherapy context remain.

Closing thoughts

The challenges posed by CSTE are not grounds for pessimism, let alone nihilism, among clinical scientists in practice or research settings, as the inferential mistakes associated with them are to some extent surmountable. Nevertheless, CSTE underscore the pressing need to inculcate humility in clinicians, researchers, and students (McFall, 1991). We are all prone to neglecting CSTE, not because of a lack of intelligence but because of inherent limitations in human information processing (Kahneman, 2011). As a consequence, all mental health professionals and consumers should be skeptical of confident proclamations of treatment breakthroughs in the absence of rigorous outcome data (Dawes, 1994; Lilienfeld et al., 2003). CSTE are potent reminders that although our intuitions are at times accurate, they can be misleading. When evaluating treatment effectiveness, our intuitions may fail to account for numerous rival hypotheses for change that are difficult or impossible to detect without the aid of finely honed research safeguards. As a consequence, CSTE highlight the inherent limits of our knowledge as applied to the individual client and should impel us to be mindful of our propensities toward overconfidence. Science, which is a systematic approach to reducing uncertainty in our inferences (McFall & Treat, 1999; O’Donohue & Lilienfeld, 2007), is ultimately our best prescription against being deceived by inadequate evidence.

Acknowledgments

The authors thank Sean Carey and Ben Johnson for their valuable assistance with compiling references.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. Note that in a case in which the client’s distress concerning her self-centeredness and hostility was itself a treatment focus, the response-shift bias would not be considered a CSTE.
2. A related cognitive error is the *subset fallacy* (Dawes, 2001). A treatment may not differ in its effects from those of a control treatment; yet on a post hoc basis, investigators may identify a subset of clients within the treatment group who displayed positive outcomes, leading them to conclude that the treatment was efficacious for that subset. Yet if there was no mean difference between the treatment and control groups, mathematically the individuals in the treatment group outside of the subset must have become worse following the intervention.
3. Nevertheless, especially among clients with marked anxiety and depression, memory biases may run in the opposite direction, predisposing them to selectively recall threatening or sad information (Coles & Heimberg, 2002; R. Elliott, Rubinsztein, Sahakian, & Dolan, 2002).

References

- Addis, M. E., & Carpenter, K. M. (2000). The treatment rationale in cognitive-behavioral therapy: Psychological mechanisms and clinical guidelines. *Cognitive Therapy and Research, 24*, 313–326.
- Alexander, F. F., & French, T. M. (1946). *Psychoanalytic therapy: Principles and application*. Oxford, England: Ronald Press.
- Alpert, J. (2012, April 22). In therapy forever? Enough already. *The New York Times*, p. SR5.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.
- American Psychological Association Task Force on Evidence-Based Practice. (2006). *Report of the 2005 Presidential Task Force on Evidence-Based Practice*. Washington, DC: American Psychological Association.
- Axson, 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. doi:10.1016/0022-1031(85)90012-5
- Bachrach, H. M., Galatzer-Levy, R., Skolnikoff, A., & Waldron, S. (1991). On the efficacy of psychoanalysis. *Journal of the American Psychoanalytic Association, 39*, 871–916.
- Baekeland, F., & Lundwall, L. (1975). Dropping out of treatment: A critical review. *Psychological Bulletin, 82*, 738–783.
- Baker, T. B., McFall, R. M., & Shoham, V. (2008). The current status and future of clinical psychology: Towards a scientifically principled approach. *Psychological Science in the Public Interest, 9*, 67–103.
- 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. doi:10.1037/0022-006X.75.6.842
- Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence*. Chichester, England: Wiley.

- Barlow, D. H. (2004). Psychological treatments. *American Psychologist*, *59*, 869–878.
- Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist-practitioner: Research and accountability in clinical and educational settings*. New York, NY: Pergamon.
- Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*, 215–220.
- Baskin, T. W., Tierney, S., Minami, T., & Wampold, B. E. (2003). Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology*, *71*, 973–979. doi:10.1037/0022-006X.71.6.973
- Beecher, H. K. (1955). The powerful placebo. *Journal of the American Medical Association*, *159*, 1602–1606.
- Belofsky, N. (2013). *Strange medicine: A shocking history of real medical practices through the ages*. New York, NY: Penguin Books.
- Beutler, L. E. (2004). The empirically supported treatments movement: A scientist-practitioner's response. *Clinical Psychology: Science and Practice*, *11*, 225–229.
- Beyerstein, B. L. (1997). Why bogus therapies seem to work. *Skeptical Inquirer*, *21*, 29–34.
- Bigby, M. (1998). Evidence-based medicine in a nutshell. *Archives of Dermatology*, *123*, 1609–1618.
- Blanco, F., Barberia, I., & Matute, H. (2014). The lack of side effects of an ineffective treatment facilitates the development of a belief in its effectiveness. *PLoS ONE*, *9*(1), e84084.
- Bloom, P. B. (1994). Is insight necessary for successful treatment? *American Journal of Clinical Hypnosis*, *36*, 172–174.
- Bloom, P., & Weisberg, D. (2007, May 18). Childhood origins of adult resistance to science. *Science*, *316*, 996–997. doi:10.1126/science.1133398
- Bohart, A. C. (2005). Evidence-based psychotherapy means evidence-informed, not evidence-driven. *Journal of Contemporary Psychotherapy*, *35*, 39–53.
- Boisvert, C. M., & Faust, D. (2002). Iatrogenic symptoms in psychotherapy: A theoretical exploration of the potential impact of labels, language and belief systems. *American Journal of Psychotherapy*, *56*, 244–259.
- Boot, W. R., Simons, D. J., Stohart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, *8*, 445–454.
- Bootzin, R. R., & Bailey, M. T. (2005). Understanding placebo, nocebo, and iatrogenic treatment effects. *Journal of Clinical Psychology*, *61*, 871–880.
- Bourne, H. (1953). The insulin myth. *Lancet*, *262*, 964–968. doi:10.1016/S0140-6736(53)91232-X
- Bowers, T. G., & Clum, G. A. (1988). Relative contribution of specific and nonspecific treatment effects: Meta-analysis of placebo-controlled behavior therapy research. *Psychological Bulletin*, *103*, 315–323.
- Braffman, W., & Kirsch, I. (1999). Imaginative suggestibility and hypnotizability: An empirical analysis. *Journal of Personality and Social Psychology*, *77*, 578–587. doi:10.1037/0022-3514.77.3.578
- Bray, J. H., Maxwell, S. E., & Howard, G. S. (1984). Methods of analysis with response-shift bias. *Educational and Psychological Measurement*, *44*, 781–804. doi:10.1177/0013164484444002
- Bridges, M. R. (2006). Activating the corrective emotional experience. *Journal of Clinical Psychology*, *62*, 551–568. doi:10.1002/jclp.20248
- Bruner, J. S. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Bull, J. P. (1959). The historical development of clinical therapeutic trials. *Journal of Chronic Diseases*, *10*, 218–248.
- Busch, A. M., Kanter, J. W., Sedivy, S. K., & Leonard, J. L. (2007). A follow-up analogue study on the effectiveness of the cognitive rationale. *Cognitive Therapy and Research*, *31*, 805–815.
- Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York, NY: Guilford Press.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand McNally.
- Carey, B. (2012, May 18). Psychiatry giant sorry for backing gay "cure." *The New York Times*. Available from <http://www.nytimes.com>
- Carlier, I. E., Voerman, A. E., & Gersons, B. R. (2000). The influence of occupational debriefing on post-traumatic stress symptomatology in traumatized police officers. *British Journal of Medical Psychology*, *73*, 87–98. doi:10.1348/000711200160327
- Carroll, K. M., Rounsaville, B. J., & Nich, C. (1994). Blind man's bluff: Effectiveness and significance of psychotherapy and pharmacotherapy blinding procedures in a clinical trial. *Journal of Consulting and Clinical Psychology*, *62*, 276–280. doi:10.1037/0022-006X.62.2.276
- Castelnuovo, G. (2010). Empirically supported treatments in psychotherapy: Towards an evidence-based or evidence-biased psychology in clinical settings? *Frontiers in Psychology*, *1*, 27–38.
- Cautin, R. L. (2008). David Shakow and schizophrenia research at Worcester State Hospital: The roots of the scientist-practitioner model. *Journal of the History of the Behavioral Sciences*, *44*, 219–237. doi:10.1002/jhbs.20312
- Chabris, C., & Simons, D. (2010). *The invisible gorilla: And other ways our intuitions deceive us*. New York, NY: Crown.
- Chadwell, B. H., & Howell, R. J. (1979). Research note: A self-report measure of spontaneous remission and psychotherapy outcome. *Journal of Community Psychology*, *7*, 69–71. doi:10.1002/1520-6629(197901)7:1<69::AID-JCOP2290070113>3.0.CO;2-6
- Chambless, D. L. (in press). Can we talk? Fostering interchange between scientists and practitioners. *Behavior Therapy*.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7–18.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716. doi:10.1146/annurev.psych.52.1.685
- Chan, A., & Altman, D. G. (2005). Identifying outcome reporting bias in randomised trials on PubMed: Review of publications and survey of authors. *British Medical Journal*, *330*, 753–759. doi:10.1136/bmj.38356.424606.8F

- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72*, 193–204. doi:10.1037/h0024670
- Chwalisz, K. (2003). Evidence-based practice: A framework for twenty-first-century scientist-practitioner training. *Counseling Psychologist, 31*, 497–528.
- Coles, M. E., & Heimberg, R. G. (2002). Memory biases in the anxiety disorders: Current status. *Clinical Psychology Review, 22*, 587–627. doi:10.1016/S0272-7358(01)00113-1
- Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology, 47*, 738–748. doi:10.1037/0022-3514.47.4.738
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field studies*. Skokie, IL: Rand McNally.
- Cooper, J. (1980). Reducing fears and increasing assertiveness: The role of dissonance reduction. *Journal of Experimental Social Psychology, 16*, 199–213. doi:10.1016/0022-1031(80)90064-5
- Cooper, J., & Axson, D. (1982). Effort justification in psychotherapy. In G. Weary & H. Mirels (Eds.), *Integrations of clinical and social psychology* (pp. 98–121). New York, NY: Oxford University Press.
- Coover, J. E., & Angell, F. (1907). General practice effect of special exercise. *American Journal of Psychology, 18*, 328–340. doi:10.2307/1412596
- Coren, S. (2003). Sensation and perception. In D. K. Freedheim (Ed.), *Handbook of psychology: History of psychology* (Vol. 1, pp. 85–108). Hoboken, NJ: Wiley.
- Coronary Drug Project Research Group. (1975). Clofibrate and niacin in coronary heart disease. *Journal of the American Medical Association, 231*, 360–381.
- Corsini, R. J., & Wedding, D. (2010). *Current psychotherapies*. Pacific Grove, CA: Wadsworth.
- Costello, E. J., & Janiszewski, S. (1990). Who gets treated? Factors associated with referral in children with psychiatric disorders. *Acta Psychiatrica Scandinavica, 81*, 523–529.
- Coyne, J. (2012). The antidepressant wars, a sequel: How the media distort findings, and do harm to patients. *PLOS Blogs*. Retrieved from <http://blogs.plos.org/mindthebrain/tag/placebo-washout/>
- Cromer, A. (1993). *Uncommon sense: The heretical nature of science*. New York, NY: Oxford University Press.
- Cross, K. P. (1977). Not can, but will college teaching be improved? *New Directions for Higher Education, 1977*(17), 1–15.
- Crumlish, N., & Kelly, B. D. (2009). How psychiatrists think. *Advances in Psychiatric Treatment, 15*, 72–79.
- Dawes, R. M. (1986). Representative thinking in clinical judgment. *Clinical Psychology Review, 6*, 425–441. doi:10.1016/0272-7358(86)90030-9
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York, NY: Free Press.
- Dawes, R. M. (2001). *Everyday irrationality: How pseudo-scientists, lunatics, and the rest of us systematically fail to think rationally*. Boulder, CO: Westview Press.
- Dehue, T. (2000). From deception trials to control reagents: The introduction of the control group about a century ago. *American Psychologist, 55*, 264–268. doi:10.1037/0003-066X.55.2.264
- Dehue, T. (2005). History of the control group. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in the behavioral science* (pp. 829–836). Chichester, England: Wiley.
- 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. doi:10.1093/clipsy/bpi022
- DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research, 14*, 469–482. doi:10.1007/BF01172968
- Diefenbach, G. J., Diefenbach, D., Baumeister, A., & West, M. (1999). Portrayal of lobotomy in the popular press: 1935–1960. *Journal of the History of the Neurosciences, 8*, 60–69. doi:10.1076/jhin.8.1.60.1766
- Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist, 65*, 21–33.
- D'Silva, K., Duggan, C., & McCarthy, L. (2004). Does treatment really make psychopaths worse? A review of the evidence. *Journal of Personality Disorders, 18*, 163–177. doi:10.1521/pedi.18.2.163.32775
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*, 69–106. doi:10.1111/j.1529-1006.2004.00018.x
- Edens, J. F., Skeem, J. L., Cruise, K. R., & Cauffman, E. (2001). Assessment of “juvenile psychopathy” and its association with violence: A critical review. *Behavioral Sciences & the Law, 19*, 53–80. doi:10.1002/bsl.425
- Elliott, R., Rubinsztein, J. S., Sahakian, B. J., & Dolan, R. J. (2002). The neural basis of mood-congruent processing biases in depression. *Archives of General Psychiatry, 59*, 597–604. doi:10.1001/archpsyc.59.7.597
- Elliott, S. A., & Brown, J. L. (2002). What are we doing to waiting list controls? *Behaviour Research and Therapy, 40*, 1047–1052. doi:10.1016/S0005-7967(01)00082-1
- Ellis, A. (2003). Helping people get better rather than merely feel better. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 21*, 169–182.
- Engel, R. J., & Schutt, R. K. (2012). *The practice of research in social work*. New York, NY: Sage.
- Ernst, E., & Resch, K. L. (1995). Concept of true and perceived placebo effects. *British Medical Journal, 311*, 551–553.
- Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16*, 319–324. doi:10.1037/h0063633
- Finocchiaro, M. (1981). Fallacies and the evaluation of reasoning. *American Philosophical Quarterly, 18*, 13–22.
- Fraenkel, J. R., & Wallen, N. E. (1993). *How to design and evaluate research in education* (2nd ed.). New York, NY: McGraw-Hill.
- Frank, J. D., & Frank, J. B. (1961). *Persuasion and healing*. Baltimore, MD: Johns Hopkins University Press.
- Frazier, P., Tennen, H., Gavian, M., Park, C., Tomich, P., & Tashiro, T. (2009). Does self-reported posttraumatic growth reflect genuine positive change? *Psychological Science, 20*, 912–919.

- Fremont, J., & Craighead, L. W. (1987). Aerobic exercise and cognitive therapy in the treatment of dysphoric moods. *Cognitive Therapy and Research, 11*, 241–251. doi:10.1007/BF01183268
- Gambrell, E. (2012). *Critical thinking in clinical practice: Improving the quality of judgments and decisions*. Hoboken, NJ: Wiley.
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association. doi:10.1037/10299-002
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89.
- Garb, H. N., Lilienfeld, S. O., & 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, NY: Routledge.
- Garfield, S. (1994). Research on client variables in psychotherapy. In A. E. Bergin & S. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 190–228). Oxford, England: Wiley.
- Geers, A. L., Weiland, P. E., Kosbab, K., Landry, S. J., & Helfer, S. G. (2005). Goal activation, expectations, and the placebo effect. *Journal of Personality and Social Psychology, 89*, 143–159.
- Gehan, E. A., & Lemak, N. A. (1994). *Statistics in medical research: Developments in clinical trials*. New York, NY: Plenum.
- Ghaemi, S. N. (2009). *A clinician's guide to statistics and epidemiology in mental health: Measuring truth and uncertainty*. Cambridge, England: Cambridge University Press.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451–482.
- Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York, NY: Free Press.
- Goldfried, M. R. (2010). The future of psychotherapy integration: Closing the gap between research and practice. *Journal of Psychotherapy Integration, 20*, 386–396.
- Green, D., & Latchford, G. (2012). Practice-based evidence. In D. Green & G. Latchford (Eds.), *Maximising the benefits of psychotherapy: A practice-based evidence approach* (pp. 87–108). Chichester, United Kingdom: Wiley-Blackwell.
- Greenwald, A. G., Spangenberg, E. R., Pratkanis, A. R., & Eskenazi, J. (1991). Double-blind tests of subliminal self-help audiotapes. *Psychological Science, 2*, 119–122. doi:10.1111/j.1467-9280.1991.tb00112.x
- Grissom, R. J. (1996). The magical number $.7 \pm .2$: 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.
- Grodstein, F., Clarkson, T. B., & Manson, J. E. (2003). Understanding the divergent data on postmenopausal hormone therapy. *New England Journal of Medicine, 348*, 645–650.
- Groopman, J. (2007). *How doctors think*. Boston, MA: Houghton Mifflin.
- 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. doi:10.1037/1076-8971.2.2.293
- Gruber, R. P. (1971). Behavior therapy: Problems in generalization. *Behavior Therapy, 2*, 361–368. doi:10.1016/S0005-7894(71)80070-9
- Hall, H. (2011). Evidence-based medicine, tooth-fairy science, and Cinderella medicine. *Skeptic, 17*(1), 4–5.
- Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392–407. doi:10.1016/S0022-1031(76)80006-6
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*, 155–163. doi:10.1002/jclp.20108
- Harrington, R., Whittaker, J., Shoebridge, P., & Campbell, F. (1998). Systematic review of efficacy of cognitive behaviour therapies in childhood and adolescent depressive disorder. *British Medical Journal, 316*, 1559–1563.
- Harrison, S. (2011, August 15). “Beyond scared straight” producer shares secrets of A&E's hit series. *Channel Guide Magazine*. Available from <http://www.channelguidemagblog.com>
- Hartman, S. E. (2009). Why do ineffective therapies seem helpful? A brief review. *Chiropractic and Osteopathy, 17*, 1–7. doi:10.1186/1746-1340-17-10
- Hathaway, S. R. (1948). Some considerations relative to nondirective counseling as therapy. *Journal of Clinical Psychology, 4*, 226–231. doi:10.1002/1097-4679(194807)4:3<226::AID-JCLP2270040303>3.0.CO;2-V
- Healy, D. (2002). *Randomized controlled trials: Evidence biased psychiatry*. Retrieved from <http://www.ahrp.org/COI/healy0802.php>
- Herbert, J. D. (2003). The science and practice of empirically supported treatments. *Behavior Modification, 27*, 412–430.
- Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. *Academic Medicine, 76*(Suppl. 10), S87–S89.
- Hofmann, S. G. (2008). Cognitive processes during fear acquisition and extinction in animals and humans: Implications for exposure therapy of anxiety disorders. *Clinical Psychology Review, 28*, 199–210.
- Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal, 319*, 670–674. doi:10.1136/bmj.319.7211.670
- Holmes, D. S. (1971). Round Robin therapy: A technique for implementing the effects of psychotherapy. *Journal of Consulting and Clinical Psychology, 37*, 324–331. doi:10.1037/h0031968
- Horney, K. (1945). *Our inner conflicts*. Oxford, England: W.W. Norton.
- Horowitz, M. J., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*, 209–218.
- Horowitz, S. (2012). New perspectives on the placebo effect: Implications for research and clinical practice. *Alternative and Complementary Therapies, 18*, 130–135.

- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*, 9–16.
- Hoshmand, L. T., & Polkinghorne, D. E. (1992). Redefining the science-practice relationship and professional training. *American Psychologist, 47*, 55–66.
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review, 4*, 93–106. doi:10.1177/0193841X8000400105
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology, 64*, 144–150. doi:10.1037/0021-9010.64.2.144
- Howard, K. I., Kopta, S., Krause, M. S., & Orlinsky, D. E. (1986). The dose–effect relationship in psychotherapy. *American Psychologist, 41*, 159–164. doi:10.1037/0003-066X.41.2.159
- Hróbjartsson, A., & Gotzsche, P. C. (2001). Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine, 344*, 1594–1602. doi:10.1056/NEJM200105243442106
- Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 131–137. doi:10.1037/0022-006X.57.1.131
- Huck, S. W., & Sandler, H. M. (1979). *Rival hypotheses: Alternative interpretations of data based conclusions*. New York, NY: Harper & Row.
- Hume, D. (1748). *Enquiry concerning human understanding*. London, England: P. F. Collier & Son.
- Hunsley, J., & Di Giulio, G. (2002). Dodo bird, phoenix, or urban legend? The question of psychotherapy equivalence. *Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work, 1*, 11–22.
- Ilardi, S. S., & Craighead, W. (1999). Rapid early response, cognitive modification, and nonspecific factors in cognitive behavior therapy for depression: A reply to Tang and DeRubeis. *Clinical Psychology: Science and Practice, 6*, 295–299.
- Ioannidis, J. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association, 294*, 218–228. doi:10.1001/jama.294.2.218
- Jacobson, N. S., & Christensen, A. (1996). Studying the effectiveness of psychotherapy: How well can clinical trials do the job? *American Psychologist, 51*, 1031–1039. doi:10.1037/0003-066X.51.10.1031
- James, F. E. (1992). Insulin treatment in psychiatry. *History of Psychiatry, 3*, 221–235.
- Jones, K. (2000). Insulin coma therapy in schizophrenia. *Journal of the Royal Society of Medicine, 93*, 147–149.
- Jopling, D. A. (2001). Placebo insight: The rationality of insight-oriented psychotherapy. *Journal of Clinical Psychology, 57*, 19–36. doi:10.1002/1097-4679(200101)57:1<19::AID-JCLP4>3.0.CO;2-Z
- Jopling, D. A. (2008). *Talking cures and placebo effects*. Oxford, England: Oxford University Press.
- Jorm, A. F., Duncan-Jones, P. P., & Scott, R. R. (1989). An analysis of the re-test artefact in longitudinal studies of psychiatric symptoms and personality. *Psychological Medicine, 19*, 487–493. doi:10.1017/S0033291700012514
- Kahneman, D. (1965). Control of spurious association and the reliability of the controlled variable. *Psychological Bulletin, 64*, 326–329. doi:10.1037/h0022529
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*, 515–526. doi:10.1037/a0016755
- Kanter, J. W., Kohlenberg, R. J., & Loftus, E. F. (2004). Experimental and psychotherapeutic demand characteristics and the cognitive therapy rationale: An analogue study. *Cognitive Therapy and Research, 28*, 229–239.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification, 1*, 427–452. doi:10.1177/014544557714001
- Kazdin, A. E. (2007). Systematic evaluation to improve the quality of patient care: From hope to hopeful. *Pragmatic Case Studies in Psychotherapy, 3*(4), 37–49.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*, 1116–1129.
- 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*, 241–250. doi:10.1016/S0740-5472(03)00021-7
- Kendall, P. C., Butcher, J. N., & Holmbeck, G. N. (1999). *Handbook of research methods in clinical psychology* (2nd ed.). Hoboken, NJ: Wiley.
- Kienle, G. S., & Kiene, H. (1997). The powerful placebo effect: Fact or fiction? *Journal of Clinical Epidemiology, 50*, 1311–1318.
- Kirsch, I. (2005). Placebo psychotherapy: Synonym or oxymoron? *Journal of Clinical Psychology, 61*, 791–803.
- Kirsch, I. (2013). The placebo effect revisited: Lessons learned to date. *Complementary Therapies in Medicine, 21*, 102–104.
- Kirsch, I., & Sapirstein, G. (1998). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention & Treatment, 1*(2), Article 0002a. doi:10.1037/1522-3736.1.1.12a
- Klein, D. F. (1998). Listening to meta-analysis but hearing bias. *Prevention & Treatment, 1*(2), Article 6c. doi:10.1037/1522-3736.1.1.16c
- Koffka, K. K. (1935). *Principles of Gestalt psychology*. Oxford, England: Harcourt, Brace.
- Koocher, G. P., Norcross, J. C., & Hill, S. S. (Eds.). (2005). *Psychologists' desk reference* (2nd ed.). New York, NY: Oxford University Press.
- 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.
- Kruger, J., Savitsky, K., & Gilovich, T. (1999). Superstition and the regression effect. *Skeptical Inquirer, 23*, 24–29.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498. doi:10.1037/0033-2909.108.3.480

- Lambert, M. J. (1976). Spontaneous remission in adult neurotic disorders: A revision and summary. *Psychological Bulletin*, 83, 107–118.
- Lambert, M. J. (2005). Early response in psychotherapy: Further evidence for the importance of common factors rather than “placebo effects.” *Journal of Clinical Psychology*, 61, 855–869.
- Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N., & Burlingame, G. M. (1994). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.1)*. Salt Lake City, UT: IHC Center for Behavioral Healthcare Efficacy.
- Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139–193). New York, NY: Wiley.
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311–328. doi:10.1037/0022-3514.32.2.311
- Lassiter, G., Geers, A. L., Munhall, P., Ploutz-Snyder, R. J., & Breitenbecher, D. L. (2002). Illusory causation: Why it occurs. *Psychological Science*, 13, 299–305. doi:10.1111/j.0956-7976.2002.x
- Laurenceau, J. P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, 27, 682–695.
- Lehrer, J. (2010, February 25). Depression's upside. *The New York Times*. Available from <http://www.nytimes.com>
- Leins, U., Goth, G., Hinterberger, T., Klinger, C., Rumpf, N., & Strehl, U. (2007). Neurofeedback for children with ADHD: A comparison of SCP and Theta/Beta protocols. *Applied Psychophysiology and Biofeedback*, 32, 73–88.
- Lenzenweger, M. F. (1999). Stability and change in personality disorder features: The longitudinal study of personality disorders. *Archives of General Psychiatry*, 56, 1009–1015. doi:10.1001/archpsyc.56.11.1009
- Leucht, S. S., Arbter, D. D., Engel, R. R., Kissling, W. W., & Davis, J. M. (2009). How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. *Molecular Psychiatry*, 14, 429–447. doi:10.1038/sj.mp.4002136
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2, 53–70. doi:10.1111/j.1745-6916.2007.00029.x
- Lilienfeld, S. O. (2014). The Dodo Bird verdict: Status in 2014. *Behavior Therapist*, 37, 91–95.
- Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, 4, 390–398. doi:10.1111/j.1745-6924.2009.01144.x
- Lilienfeld, S. O., Lohr, J. M., & Olatunji, B. O. (2008). Encouraging students to think critically about psychotherapy: Overcoming naïve realism. In D. S. Dunn, J. S. Halonen, & R. A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 267–271). Chichester, United Kingdom: Wiley-Blackwell. doi:10.1002/9781444305173.ch26
- Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2003). *Science and pseudoscience in clinical psychology*. New York, NY: 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, 883–890.
- Litz, B. T., Gray, M. J., Bryant, R. A., & Adler, A. B. (2002). Early intervention for trauma: Current status and future directions. *Clinical Psychology: Science and Practice*, 9, 112–134. doi:10.1093/clipsy/9.2.112
- Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist*, 48, 518–537. doi:10.1037/0003-066X.48.5.518
- Loftus, L. S., & Arnold, W. N. (1991). Vincent van Gogh's illness: Acute intermittent porphyria? *British Medical Journal*, 303, 1589–1591.
- Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clinical Psychology Review*, 27, 173–187.
- Loranger, A. W., Lenzenweger, M. F., Gartner, A. F., & Susman, V. L. (1991). Trait-state artifacts and the diagnosis of personality disorders. *Archives of General Psychiatry*, 48, 720–728. doi:10.1001/archpsyc.1991.01810320044007
- 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. doi:10.1001/archpsyc.1975.01760260059004
- Luty, S. E., Carter, J. D., McKenzie, J. M., Rae, A. M., Frampton, C. M., Mulder, R. T., & Joyce, P. R. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *British Journal of Psychiatry*, 190, 496–502.
- Lynch, T. R., Chapman, A. L., Rosenthal, M., Kuo, J. R., & Linehan, M. M. (2006). Mechanisms of change in dialectical behavior therapy: Theoretical and empirical observations. *Journal of Clinical Psychology*, 62, 459–480. doi:10.1002/jclp.20243
- Lynn, S. J., & Nash, M. R. (1994). Truth in memory: Ramifications for psychotherapy and hypnotherapy. *American Journal of Clinical Hypnosis*, 36(3), 194–208.
- Magaret, A. (1950). Generalization in successful psychotherapy. *Journal of Consulting Psychology*, 14, 64–70. doi:10.1037/h0053633
- Manzi, J. (2012). *Uncontrolled: The surprising payoff of trial-and-error in business, politics and society*. New York, NY: Basic Books.
- Marino, L., & Lilienfeld, S. O. (2007). Dolphin-assisted therapy: More flawed data and more flawed conclusions. *Anthrozoös*, 20, 239–249. doi:10.2752/089279307X224782
- Markin, R. D., & Kivlighan, D. (2007). Bias in psychotherapist ratings of client transference and insight. *Psychotherapy: Theory, Research, Practice, Training*, 44, 300–315. doi:10.1037/0033-3204.44.3.300
- Marmar, C. R. (1990). Psychotherapy process research: Progress, dilemmas, and future directions. *Journal of Consulting and Clinical Psychology*, 58, 265–272.
- Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *The Quarterly Journal of Experimental Psychology*, 65, 2271–2279.

- Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology*, *102*, 392–405.
- Mausbach, B. T., Moore, R., Roesch, S., Cardenas, V., & Patterson, T. L. (2010). The relationship between homework compliance and therapy outcomes: An updated meta-analysis. *Cognitive Therapy and Research*, *34*, 429–438. doi:10.1007/s10608-010-9297-z
- Mayou, R. A., Ehlers, A. A., & Hobbs, M. M. (2000). Psychological debriefing for road traffic accident victims: Three-year follow-up of a randomised controlled trial. *British Journal of Psychiatry*, *176*, 589–593. doi:10.1192/bjp.176.6.589
- Mayr, E. (1981, October 30). Biological classification: Toward a synthesis of opposing methodologies. *Science*, *214*, 510–516.
- McArthur, L. Z. (1980). Illusory causation and illusory correlation: Two epistemological accounts. *Personality and Social Psychology Bulletin*, *6*, 507–519. doi:10.1177/014616728064003
- McArthur, L. Z., & Solomon, L. K. (1978). Perceptions of an aggressive encounter as a function of the victim's salience and the perceiver's arousal. *Journal of Personality and Social Psychology*, *36*, 1278–1290. doi:10.1037/0022-3514.36.11.1278
- McCall, R. B. (1977). Challenges to a science of developmental psychology. *Child Development*, *48*, 333–344. doi:10.2307/1128626
- McCauley, R. (2011). *Why religion is natural and science is not*. Oxford, England: Oxford University Press.
- McCullough, J. P. (2000). *Treatment for chronic depression: Cognitive Behavioral Analysis System of Psychotherapy (CBASP)*. New York, NY: Guilford Press.
- McDonald, C. J., Mazzuca, S. A., & McCabe, G. P. (1983). How much of the placebo "effect" is really statistical regression? *Statistics in Medicine*, *2*, 417–427.
- McFall, R. M. (1991). Manifesto for a science of clinical psychology. *The Clinical Psychologist*, *44*(6), 75–88.
- McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology*, *50*, 215–241. doi:10.1146/annurev.psych.50.1.215
- McFarland, C., & Alvaro, C. (2000). The impact of motivation on temporal comparisons: Coping with traumatic events by perceiving personal growth. *Journal of Personality and Social Psychology*, *79*, 327–343.
- McHugh, P. M. (1994). Psychotherapy awry. *American Scholar*, *63*, 17–30.
- McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*, *65*, 73–84.
- McLeod, J. (2001). An administratively created reality: Some problems with the use of self-report questionnaire measures of adjustment in counselling/psychotherapy outcome research. *Counselling & Psychotherapy Research*, *1*, 215–226.
- 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. doi:10.1111/1529-1006.01421
- McReynolds, W. T., & Tori, C. (1972). A further assessment of attention-placebo effects and demand characteristics in studies of systematic desensitization. *Journal of Consulting and Clinical Psychology*, *38*, 261–264.
- Meehl, P. E. (1955). Psychotherapy. *Annual Review of Psychology*, *63*, 57–378. doi:10.1146/annurev.ps.06.020155.002041
- Messer, S. B., & Wampold, B. E. (2002). Let's face facts: Common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice*, *9*, 21–25.
- Meyer, R. G., & Deitsch, S. E. (1996). *The clinician's handbook: Integrated diagnostics, assessment, and intervention in adult and adolescent psychopathology* (4th ed.). Needham Heights, MA: Allyn & Bacon.
- Michotte, A. (1945). *The perception of causality*. New York, NY: Basic Books.
- Mintz, J., Drake, R. E., & Crits-Christoph, P. (1996). Efficacy and effectiveness of psychotherapy: Two paradigms, one science. *American Psychologist*, *51*, 1084–1085. doi:10.1037/0003-066X.51.10.1084
- Morgado, A., Smith, M., Lecrubier, Y., & Widlocher, D. (1991). Depressed subjects unwittingly overreport poor social adjustment which they reappraise when recovered. *Journal of Nervous and Mental Disease*, *179*, 614–619.
- Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry*, *34*, 289–306.
- Morrison, J. R. (1997). *When psychological problems mask medical disorders: A guide for psychotherapists*. New York, NY: Guilford Press.
- Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2003). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? In A. E. Kazdin (Ed.), *Methodological issues & strategies in clinical research* (3rd ed., pp. 505–546). Washington, DC: American Psychological Association.
- Nichols, M. P., & Efran, J. S. (1985). Catharsis in psychotherapy: A new perspective. *Psychotherapy: Theory, Research, Practice, Training*, *22*, 46–58.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. doi:10.1037/1089-2680.2.2.175
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Nock, M. K., Michel, B. D., & Photos, V. I. (2007). Single-case research designs. In D. McKay (Ed.), *Handbook of research methods in abnormal and clinical psychology* (pp. 337–350). New York, NY: Sage.
- Norcross, J. C., Beutler, L. E., & Levant, R. F. (2007). *Evidence-based practices in mental health*. Washington, DC: American Psychological Association.
- Norcross, J. C., & Lambert, M. J. (2006). The therapy relationship. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 208–218). Washington, DC: American Psychological Association.

- Novella, S. E. (2008). The placebo effect. *Science-Based Medicine*. Retrieved from <http://www.sciencebasedmedicine.org/index.php/the-placebo-effect/>
- Novella, S. E. (2010). The poor, misunderstood placebo. *Skeptical Inquirer*, 34(6), 33–34.
- O'Donohue, W., & Lilienfeld, S. O. (2007). The epistemological and ethical dimension of clinical science. In T. A. Treat, R. R. Bootzin, & T. B. Baker (Eds.), *Psychological clinical science: Papers in honor of Richard M. McFall* (pp. 29–52). New York, NY: Psychology Press.
- Offit, P. A. (2010). *Autism's false prophets: Bad science, risky medicine, and the search for a cure*. New York, NY: Columbia University Press.
- O'Leary, K., & Borkovec, T. D. (1978). Conceptual, methodological, and ethical problems of placebo groups in psychotherapy research. *American Psychologist*, 33, 821–830. doi:10.1037/0003-066X.33.9.821
- Orlinsky, D. E., Rønnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 307–390). New York, NY: Wiley.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–783. doi:10.1037/h0043424
- Pace, T. W., Negi, L., Adame, D. D., Cole, S. P., Sivilli, T. I., Brown, T. D., . . . Raison, C. L. (2009). Effect of compassion meditation on neuroendocrine, innate immune and behavioral responses to psychosocial stress. *Psychoneuroendocrinology*, 34, 87–98. doi:10.1016/j.psyneuen.2008.08.011
- Paul, G. L. (1966). *Insight vs. desensitization in psychotherapy: an experiment in anxiety reduction*. Stanford, CA: Stanford University Press.
- Penedo, F. J., & Dahn, J. R. (2005). Exercise and well-being: A review of mental and physical health benefits associated with physical activity. *Current Opinion in Psychiatry*, 18, 189–193. doi:10.1097/00001504-200503000-00013
- Pennebaker, J. W., & Skelton, J. A. (1981). Selective monitoring of physical sensations. *Journal of Personality and Social Psychology*, 41, 213–223. doi:10.1037/0022-3514.41.2.213
- Petitti, D. B. (1994). *Meta-analysis, decision analysis and cost-effectiveness analysis*. New York, NY: Oxford University Press.
- Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2005). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency. *The Scientific Review of Mental Health Practice*, 4, 48–54.
- Posternak, M. A., & Zimmerman, M. (2000). Short-term spontaneous improvement in depressed outpatients. *Journal of Nervous and Mental Disease*, 188, 799–804.
- Poulsen, S., Lunn, S., Daniel, S. I., Folke, S., Mathiesen, B. B., Katznelson, H., & Fairburn, C. G. (2014). A randomized controlled trial of psychoanalytic psychotherapy or cognitive-behavioral therapy for bulimia nervosa. *American Journal of Psychiatry*, 171, 109–116.
- Prasad, V., Cifu, A., & Ioannidis, J. P. (2012). Reversals of established medical practices. *Journal of the American Medical Association*, 307, 37–38.
- Rachman, S. J. (1973). The effects of psychological treatments. In H. Eysenck (Ed.), *Handbook of abnormal psychology* (pp. 805–819). New York, NY: Basic Books.
- Rice, M. E., Harris, G. T., & Cormier, C. A. (1992). An evaluation of a maximum security therapeutic community for psychopaths and other mentally disordered offenders. *Law and Human Behavior*, 16, 399–412. doi:10.1007/BF02352266
- Rickels, K., Lipman, R. S., Fisher, S., Park, L. C., & Uhlenhuth, E. H. (1970). Is a double-blind clinical trial really double-blind? A report of doctors' medication guesses. *Psychopharmacologia*, 16, 329–336.
- Ritschel, L. A. (2005). Reconciling the rift: Improving clinical psychology graduate training in the twenty-first century. *Journal of Clinical Psychology*, 61, 1111–1114. doi:10.1002/jclp.20147
- Robins, R. W., Fraley, R., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, 69, 617–640. doi:10.1111/1467-6494.694157
- Robins, R. W., Nofle, E. E., Trzesniewski, K. H., & Roberts, B. W. (2005). Do people know how their personality has changed? Correlates of perceived and actual personality change in young adulthood. *Journal of Personality*, 73, 489–521. doi:10.1111/j.1467-6494.2005.00317.x
- Romanczyk, R. G., Arnstein, L., Soorya, L. V., & Gillis, J. (2003). The myriad of controversial treatments for autism: A critical evaluation of efficacy. In S. O. Lilienfeld, S. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 363–395). New York, NY: Guilford Press.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641. doi:10.1037/0033-2909.86.3.638
- Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 103–135). Hillsdale, NJ: Erlbaum.
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, 96, 341–357. doi:10.1037/0033-295X.96.2.341
- Roth, A., & Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research*. New York, NY: Guilford Press.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312, 71–72.
- Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the 20th century*. New York, NY: W.H. Freeman.
- Samuel, D. B., Hopwood, C. J., Ansell, E. B., Morey, L. C., Sanislow, C. A., Markowitz, J. C., . . . Grilo, C. M. (2011). Comparing the temporal stability of self-report and interview assessed personality disorder. *Journal of Abnormal Psychology*, 120, 670–680. doi:10.1037/a0022647
- Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception*. Oxford, England: Bobbs-Merrill.
- Seligman, M. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist*, 50, 965–974. doi:10.1037/0003-066X.50.12.965

- Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., & Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy*, *47*, 902–909.
- Shapiro, A. K., & Shapiro, E. (1997). The placebo: Is it much ado about nothing? In A. Harrington (Ed.), *The placebo effect: An interdisciplinary exploration* (pp. 12–36). Cambridge, MA: Harvard University Press.
- Shea, M., Edelen, M. O., Pinto, A. A., Yen, S. S., Gunderson, J. G., Skodol, A. E., & . . . Morey, L. C. (2009). Improvement in borderline personality disorder in relationship to age. *Acta Psychiatrica Scandinavica*, *119*, 143–148. doi:10.1111/j.1600-0447.2008.01274.x
- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, *65*, 98–109. doi:10.1037/a0018378
- Slovan, S. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University Press.
- Smith, D. B. (2009, September). The doctor is in. *The American Scholar*, *78*(4). Available from <http://theamericanscholar.org>
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.
- Solomon, D. A., Leon, A. C., Coryell, W. H., Endicott, J., Li, C., Fiedorowicz, J. G., . . . Keller, M. B. (2010). Longitudinal course of bipolar I disorder: Duration of mood episodes. *Archives of General Psychiatry*, *67*, 339–347. doi:10.1001/archgenpsychiatry.2010.15
- Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., & Rush, J. D. (2009). The Meta-Analysis of Clinical Judgment Project: Effects of experience on judgment accuracy. *The Counseling Psychologist*, *37*, 350–399.
- Spitzer, R. L. (2003). Can some gay men and lesbians change their sexual orientation? 200 participants reporting a change from homosexual to heterosexual orientation. *Archives of Sexual Behavior*, *32*, 403–417. doi:10.1023/A:1025647527010
- Spitzer, R. L. (2012). Spitzer reassesses his 2003 study of reparative therapy of homosexuality. *Archives of Sexual Behavior*, *41*, 747. doi:10.1007/s10508-012-9966-y
- 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. doi:10.1002/jclp.20373
- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*, 672–695.
- Steer, R. Y., & Ritschel, L. A. (2010). Placebo. In I. Weiner & W. E. Craighead (Eds.), *Corsini encyclopedia of psychology* (pp. 1252–1254). New York, NY: Wiley.
- Stewart, R. E., & Chambless, D. L. (2007). Does psychotherapy research inform treatment decisions in private practice? *Journal of Clinical Psychology*, *63*, 267–281. doi:10.1002/jclp.20347
- Stewart, R. E., Chambless, D. L., & Baron, J. (2011). Theoretical and practical barriers to practitioners' willingness to seek training in empirically supported treatments. *Journal of Clinical Psychology*, *68*, 8–23.
- Straus, S. E., Tetroe, J., Graham, I. D., Zwarenstein, M., Bhattacharyya, O., & Shepperd, S. (2010). Monitoring use of knowledge and evaluating outcomes. *Canadian Medical Association Journal*, *182*, e94–e98. doi:10.1503/cmaj.081335
- Stricker, G. (2003). The many faces of self-disclosure. *Journal of Clinical Psychology*, *59*, 623–630.
- Stricker, G., & Trierweiler, S. J. (1995). The local clinical scientist: A bridge between science and practice. *American Psychologist*, *50*, 995–1002. doi:10.1037/0003-066X.50.12.995
- Suls, J., & Howren, M. B. (2012). Understanding the physical-symptom experience: The distinctive contributions of anxiety and depression. *Current Directions in Psychological Science*, *21*, 129–134.
- Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*, 547–559. doi:10.1037/a0028226
- Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. New York, NY: Random House.
- 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.
- Taubes, G. (2007, September 16). Do we really know what makes us healthy? *The New York Times*. Available from <http://www.dscience.net>
- Tavris, C. (2003). The widening scientist–practitioner gap. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. ix–xvii). New York, NY: Guilford Press.
- Taylor, S. E., & Fiske, S. T. (1975). Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, *32*, 439–445. doi:10.1037/h0077095
- Taylor, S. E., & Fiske, S. T. (1978). Saliency, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 249–288). New York, NY: Academic Press.
- Tehrani, E. E., Krussel, J. J., Borg, L. L., & Munk-Jørgensen, P. P. (1996). Dropping out of psychiatric treatment: A prospective study of a first-admission cohort. *Acta Psychiatrica Scandinavica*, *94*, 266–271. doi:10.1111/j.1600-0447.1996.tb09859.x
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 Restructured Clinical (RC) scales: Development, validation, and interpretation*. Minneapolis: University of Minnesota Press.
- Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.
- Thompson, S. C. (1999). Illusions of control: How we overestimate our personal influence. *Current Directions in Psychological Science*, *8*, 187–190. doi:10.1111/1467-8721.00044
- Tolin, D. F. (2010). Is cognitive-behavioral therapy more effective than other therapies?: A meta-analytic review. *Clinical Psychology Review*, *30*, 710–720. doi:10.1016/j.cpr.2010.05.003

- Tracey, T. J., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69*, 218–229.
- Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131. doi:10.1126/science.185.4157.1124
- Valenstein, E. S. (1986). *Great and desperate cures: The rise and decline of psychosurgery and other radical treatments for mental illness*. New York, NY: Basic Books.
- Wachtel, P. L. (1987). *Action and insight*. New York, NY: Guilford Press.
- Wachtel, P. L. (2010). Beyond “ESTs”: Problematic assumptions in the pursuit of evidence-based practice. *Psychoanalytic Psychology, 27*, 251–263.
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110*, 639–644. doi:10.2466/02.07.17.PR0.110.2.639-644
- Walker, W., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant – and memory helps to keep it that way! *Review of General Psychology, 7*, 203–210. doi:10.1037/1089-2680.7.2.203
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Hillsdale, NJ: Erlbaum.
- Wampold, B. E., & DeFife, J. (2010, February 6). How psychotherapy works. *Psychology Today*. Retrieved from <http://www.psychologytoday.com/blog/the-shrink-tank/201002/how-psychotherapy-works>
- 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. doi:10.1037/0033-2909.122.3.203
- Wechsler, M. E., Kelley, J. M., Boyd, I. O., Dutille, S., Marigowda, G., Kirsch, I., . . . Kaptchuk, T. J. (2011). Active albuterol or placebo, sham acupuncture, or no intervention in asthma. *New England Journal of Medicine, 365*, 119–126.
- Weiss, B., & Weisz, J. R. (1990). The impact of methodological factors on child psychotherapy outcome research: A meta-analysis for researchers. *Journal of Abnormal Child Psychology, 18*, 639–670. doi:10.1007/BF01342752
- Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*, 450–458.
- Westen, D. (1988). Transference and information processing. *Clinical Psychology Review, 8*, 161–179.
- Westen, D., & Bradley, R. (2005). Empirically supported complexity: Rethinking evidence-based practice in psychotherapy. *Current Directions in Psychological Science, 14*, 266–271. doi:10.1111/j.0963-7214.2005.00378.x
- 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. doi:10.1037/0033-2909.130.4.631
- Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2005). EBP≠EST: Reply to Crits-Christoph et al. (2005) and Weisz et al. (2005). *Psychological Bulletin, 131*, 427–433.
- Westen, D., & Weinberger, J. (2005). In praise of clinical judgment: Meehl's forgotten legacy. *Journal of Clinical Psychology, 61*, 1257–1276. doi:10.1002/jclp.20181
- Whisman, M. A., Miller, I. W., Norman, W. H., & Keitner, G. I. (1991). Cognitive therapy with depressed inpatients: Specific effects on dysfunctional cognitions. *Journal of Consulting and Clinical Psychology, 59*, 282–288.
- Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality and Social Psychology, 80*, 572–584. doi:10.1037/0022-3514.80.4.572
- Wilson, G. T. (1999). Rapid response to cognitive behavior therapy. *Clinical Psychology: Science and Practice, 6*, 289–292.
- Wilson, T. D. (2011). *Redirect: The surprising new science of psychological change*. Boston, MA: Little, Brown.
- Wolpert, L. (1992). *The unnatural nature of science*. Cambridge, MA: Harvard University Press.
- Zamansky, H. S., Scharf, B., & Brightbill, R. (1964). The effect of expectancy for hypnosis on prehypnotic performance. *Journal of Personality, 32*, 236–248. doi:10.1111/j.1467-6494.1964.tb01338.x
- Zimmerman, M., & Mattia, J. I. (1999). Is posttraumatic stress disorder underdiagnosed in routine clinical settings? *Journal of Nervous and Mental Disease, 187*, 420–428.