

Distinguishing Scientific From Pseudoscientific Psychotherapies: Evaluating the Role of Theoretical Plausibility, With a Little Help From Reverend Bayes

Scott O. Lilienfeld, Emory University

In their stimulating article, David and Montgomery (2011) underscore the importance of identifying pseudoscientific and potentially harmful psychotherapies, and observe that the criteria for empirically supported therapies neglect to take into account the theoretical support for treatment packages. They propose a framework for evidence-based therapies that includes empirical support for both the efficacy of the treatment and the theory that spawned it. I discuss several challenges to this and similar frameworks, including (a) the absence of convincing data regarding how most current therapies work, (b) the underdetermination of theory by data, making it difficult to strongly corroborate models of therapeutic change mechanisms using only a small number of studies, and (c) the risk of false negatives, viz., efficacious treatments derived from nonexistent or flawed theories. As a friendly amendment, I propose a Bayesian alternative to David and Montgomery's classification that incorporates the role of theoretical plausibility while minimizing the aforementioned difficulties.

Key words: Bayes theorem, empirically supported therapies, exposure, nonspecific effects, pseudoscience, psychotherapy. [*Clin Psychol Sci Prac* 18: 105–112, 2011]

With well over 500 psychotherapies available to mental health consumers (Eisner, 2000), professionals and would-be therapy clients require some means of sorting the wheat from the chaff. This need is pressing because the once widespread assumption that psycho-

therapy can only help, not harm, is increasingly being called into question. Controlled data suggest that certain therapies, such as crisis debriefing for traumatized individuals, Scared Straight interventions for troubled adolescents, and dissociative identity disorder-oriented therapy, sometimes lead to symptom worsening (Boisvert, 2010; Dimidjian & Hollon, 2010; Lilienfeld, 2007). Moreover, despite scattered continued claims to the contrary (Shedler, 2010; Wampold, 2001), data convincingly refute the “dodo bird verdict” of complete therapeutic equivalence. For example, behavioral therapies are more efficacious than nonbehavioral therapies for at least some anxiety disorders and childhood disorders (Chambless & Ollendick, 2001; Hunsley & DiGuilio, 2002).

In a stimulating article, David and Montgomery (2011) propose a novel framework for distinguishing scientific from pseudoscientific interventions—those that mimic the style of science but lack its substance. The proponents of pseudoscientific treatments frequently engage in maneuvers that shield their favored interventions from rigorous scrutiny, such as repeated invocation of ad hoc hypotheses (escape hatches or loopholes) to avoid falsification, evasion of peer review, emphasis on supportive rather than unsupportive evidence, absence of self-correction, and embrace of holism, viz., the claim that individual elements of a claim cannot be tested in isolation (Lilienfeld, 1998; Ruscio, 2006). In many cases, pseudosciences claim the mantle of science but refuse to play by its consensual rules.

As David and Montgomery (2011) note, classificatory efforts to place the field of psychotherapy on firmer scientific footing, such as the American Psychological Association Division 12 criteria for empirically supported therapies (ESTs), neglect to consider the research support for the theoretical mechanisms of interventions. By focusing solely on evidence for efficacious outcomes and ignoring evidence for the theoretical underpinnings of treatments, the EST criteria overlook a crucial piece of the evidentiary picture. As a consequence, David and Montgomery observe, these and other criteria for ESTs inadvertently open the floodgates for unscientific and pseudoscientific interventions to be considered empirically supported. This is especially likely, it is worth noting, because so many psychological interventions draw on nonspecific factors (e.g., placebo effects,

Address correspondence to Scott O. Lilienfeld, Ph.D., Department of Psychology, Room 473, Psychology and Interdisciplinary Sciences, Emory University, 36 Eagle Row, Atlanta, GA 30322. E-mail: slilien@emory.edu.

attention from an empathic professional, and effort justification) that can lead to improvement (Lohr, Olatunji, Parker, & DeMaio, 2005). Hence, even interventions that are largely or entirely bogus from a theoretical perspective may perform better than a wait-list control condition across multiple independent trials, and therefore satisfy the EST criteria for efficacy. For example, some have argued that energy therapies, such as thought field therapy and emotional freedom technique, now satisfy the Division 12 criteria for efficacious or at least probably efficacious interventions (Feinstein, 2008). The merits of this claim notwithstanding (e.g., McCaslin, 2009; Pignotti & Thyer, 2009), it is indeed plausible that, given their incorporation of nonspecific influences, at least some energy therapies would outperform no treatment in randomized controlled trials (RCTs) and thereby fulfill EST criteria. Yet, energy therapies are premised on the assumption that psychological problems are caused by blockages in invisible energy fields, which have never been shown to exist—and that are scientifically implausible (Saravi, 1999).

A further shortcoming of current EST lists, alluded to by David and Montgomery, is their underemphasis on distinguishing specific from nonspecific treatment components (Lohr, Lilienfeld, Tolin, & Herbert, 1999; Lohr et al., 2005). To their credit, Chambless and Hollon (1998) created a category for “efficacious and specific” interventions, reserved for treatments that have been shown not only to be efficacious in independent trials but also to outperform either a placebo or plausible alternative intervention. Nevertheless, even this category does not help to identify the ingredients of the treatment that are genuinely efficacious; this task requires systematic dismantling designs that afford strong tests of the efficacy of separable treatment components (Borkovec, 1985). As a consequence, even the efficacious and specific category may include treatments premised on dubious or even patently false theoretical assumptions. For example, eye movement desensitization and reprocessing (EMDR) appears to be more efficacious than no treatment and perhaps relaxation-based interventions (but not exposure-based therapies; Davidson & Parker, 2001). Nevertheless, the assorted theoretical rationales posited for EMDR, such as hemispheric synchronization, accelerated information processing, or simulation of

rapid eye movement sleep, have minimal support (Herbert et al., 2000).

To address the limitations of extant EST criteria and lists, David and Montgomery propose a nine-level framework for evidence-based psychotherapies that incorporates both (a) research data on treatment outcome, as in current EST criteria, and (b) research data on the theoretical mechanisms for treatments (see also Rosen & Davison, 2003). They are to be congratulated for addressing the important problem of pseudoscientific interventions, and their innovative proposal is well worth considering. In the remainder of my commentary, I consider several challenges to David and Montgomery’s classification scheme and propose an alternative conceptualization of the problem they are attempting to address.

LACK OF KNOWLEDGE REGARDING TREATMENT MECHANISMS

David and Montgomery’s scheme contains a category for treatments whose underlying theory is “well supported,” those in which the mechanisms of change have been supported in rigorous investigations. Yet, it would not be overstating the case to say that with precious few exceptions (e.g., applied tension for blood/injection/injury phobia; Ost & Sterner, 1987), *no* theoretical mechanisms of therapeutic change are well supported by research. As Kazdin (2009) noted, “with isolated exceptions, we do not know why or how therapies achieve therapeutic change” (p. 418; see also Murphy, Cooper, Hollon, & Fairburn, 2009).

This is even true for well-established behavior therapies that rely on systematic exposure to anxiety-provoking stimuli, like systematic desensitization and exposure and response prevention. There is no general agreement—or unambiguous evidence—regarding whether exposure-based treatments work by habituation, counterconditioning, extinction, changes in expectancies, alterations in connectionist networks, or some alternative mechanism (Tryon, 2005). Nor do we know whether several of these mechanisms could be describing the same phenomenon, but at different levels of explanation; for example, alterations in expectancies might reflect changes in implicit networks. Perhaps even greater ambiguity exists for the mechanism of action of cognitive therapies. Proponents (e.g., Beck,

1991) maintain that they work by altering cognitions, but the evidence for this claim is equivocal at best. For example, dismantling designs suggest that the behavioral elements of cognitive therapy, not cognitive restructuring, may explain much of its efficacy (Dimidjian et al., 2006), and other work suggests that most of the symptom reduction in cognitive therapy occurs in the first several weeks of treatment, before formal cognitive interventions have been introduced (Ilardi & Craighead, 1994). The same lack of understanding of treatment mechanisms, incidentally, holds for many somatic interventions. For example, although electroconvulsive therapy (ECT) is consistently more efficacious than control procedures (e.g., sham ECT) for treatment-refractory depression, it is not clear whether it works by means of changes in monoamine levels, changes in neural growth factors, changes in endogenous opioids, anticonvulsant effects, or still unknown processes (Taylor, 2007). The lack of clear-cut research evidence for the mechanism of action of psychological interventions poses a challenge to the David and Montgomery framework because it raises the question of whether any interventions can presently be regarded as theoretically well supported.

THE UNDERDETERMINATION OF THEORY BY EVIDENCE

To maintain consistency with the EST criteria for treatment outcome evidence, David and Montgomery propose that a “theory is well supported, within this framework, if it has been empirically validated in at least two rigorous studies, by two different investigators or investigating teams” (2011, p. 92). Here, they admirably include tests of *specific efficacy*, which include dismantling designs and mediator designs, among others, that help investigators ascertain which specific components of a treatment are efficacious.

Nevertheless, we can question whether one should use parallel criteria—in this case, two independent investigations—to ascertain both theoretical support and therapeutic efficacy. The principle of the underdetermination of theory by evidence (Quine, 1951; Stanford, 2009) reminds us that the same set of data is typically consistent with a broad array of competing theoretical explanations. Strong corroboration of a theory, including a theory of the mechanisms of action of a treatment, typically requires the accumulation of

a great deal of evidence from multiple sources. Even then, the theory can at best be regarded only as corroborated rather than confirmed (“proven”).

Ideally, science progresses by subjecting theories to multiple “risky tests,” those that place theories at grave risk of refutation (Popper, 1959). Yet risky tests are difficult to formulate in many domains of psychology (Meehl, 1978), including psychotherapy process and outcome research, largely because the substantive theory of interest is invariably conjoined with one or more auxiliary hypotheses that are detached from the substantive theory but that are needed to test it (e.g., in psychotherapy outcome research, these include hypotheses regarding investigator allegiance effects, appropriateness of the sample, treatment integrity, training of therapists, adequacy of randomization, controls for demand characteristics, and so on). As a consequence, if the data from a given study that bear on our theory are negative, it is rarely clear to what to attribute it—the substantive theory in question, one or more auxiliary hypotheses, or some conjunction of both.

Moreover, tests of the mechanisms of psychotherapies are rarely risky. As Kazdin and Nock (2003) argued persuasively, most statistical tests of mediation provide weak tests of therapeutic mechanisms, largely because many variables that do not genuinely account for the efficacy of a therapy may nevertheless pass tests of mediation. As a consequence, “drawing inferences about a mediator requires convergence of multiple criteria that act in concert” and “a sequence of studies” (Kazdin, 2009, p. 419) that provide a compelling case that a statistical mediator actually reflects the mechanism of change. The criteria for establishing a mechanism (see Kazdin, 2009, p. 420) include strength of association (is the hypothesized mediator large enough in magnitude to account for the outcome?), consistency of association across investigations (is the hypothesized mediator replicated in trials by independent researchers?), specificity (does only the hypothesized variable mediate the relation between therapy and outcome, or do other factors mediate it as well?), temporal ordering (does the hypothesized mediator change before the outcome changes?), experimental manipulation (does altering the hypothesized mediator in an RCT produce changes in the outcome?), and gradient (do stronger “doses” of the mediator relate to greater change in

outcomes?). The key point is that a few studies are rarely sufficient to establish the mechanism of a treatment with any degree of certainty.

THE RISK OF FALSE NEGATIVES

Presumably, we can all agree that client outcomes are our paramount consideration, and that any classification of evidence-based psychotherapies must place this criterion front and center. One challenge to David and Montgomery's classificatory scheme—and to others that explicitly incorporate evidence for theoretical mechanisms—is the risk of false negatives, viz., treatments that are clearly efficacious but that are based on absent or weak theoretical evidence.

Imagine that a therapist stumbles upon what appears to be a highly effective new treatment for Disorder X. He happens upon it by sheer luck with a few clients and has no idea why it seems to work. After some pilot testing, he and his colleagues examine this intervention in several RCTs and find that it is far more efficacious than any extant treatment for Disorder X. Intrigued, several independent investigators test it too, and sure enough, it is both highly efficacious and much more efficacious than any other treatments for that disorder. The researchers and clinicians are, however, all baffled by why the treatment works. According to the David and Montgomery scheme, this intervention would be classified only within Category II (Well-Supported Treatment Package, No Data for Theory), even though it is undeniably efficacious. Indeed, the efficacy data for this mysterious treatment could well be considerably stronger (e.g., much larger effect sizes, far more independent replications) than for interventions classified within Category I (Well-Supported Treatment Package, Well-supported Data for Theory).

Pushing the argument a step further, let us imagine another therapist who happens upon an extremely efficacious treatment for Disorder Y—one that is corroborated in multiple independent RCTs—but who holds an entirely erroneous and poorly supported theoretical explanation (e.g., the treatment works by synchronizing brain waves in the patient's left and right hemispheres) for the intervention's efficacy. According to David and Montgomery, this treatment would fall under Category III (Well-Supported Treatment Package, Strong Contradictory Evidence for Theory), even

though it again may be far more efficacious than any known treatment for Disorder Y.

So any scheme that incorporates theoretical support as a gatekeeper could inadvertently deprive clients of the most efficacious treatments. This could occur if mental health professionals assume that higher categories in the hierarchy are necessarily superior to those lower in the hierarchy.

THE ROLE OF TREATMENT PLAUSIBILITY: A BAYESIAN ALTERNATIVE

Despite my reservations with David and Montgomery's innovative classification scheme, I suspect that their intuitions are correct and touch on a crucial truth: theoretical plausibility should matter when evaluating the evidential support for psychotherapies. For example, when EMDR was classified as "probably efficacious" for civilian post-traumatic stress disorder in an influential article on ESTs (Chambless & Hollon, 1998), it provoked an outcry from many skeptics, myself included. Yet to a large extent, the outcry was misplaced because EMDR almost certainly fulfilled the EST criteria for probable efficacy. Nevertheless, the skeptics were on to something. Although EMDR was indeed shown in several studies to be more efficacious than no treatment, there was no compelling or even especially suggestive evidence that the eye movements or other ostensibly theoretically distinctive features of the treatment mattered (a conclusion later confirmed by meta-analyses; see Davidson & Parker, 2001). Nor was there a shred of plausible evidence for the theoretical mechanisms (e.g., accelerated neural information processing) put forth for the treatment by EMDR's developers (e.g., Shapiro, 1998). EMDR is theoretically dubious at best (many used far harsher language), skeptics groused; why should it be considered an EST?

In reality, I doubt that many skeptics doubted the controlled evidence that EMDR worked better than nothing. Instead, I suspect that most doubted whether EMDR possessed *specific efficacy*; that is, they doubted whether any of the ostensibly distinctive components of the treatment, such as eye movements, do much of anything. Put a bit differently, the skeptics were contending, perhaps implicitly, that "given the low theoretical plausibility of EMDR's therapeutic mechanisms, several findings indicating that it is more efficacious

than no treatment do little to boost my belief that EMDR adds anything to what's out there."

In essence, the skeptics were acting as implicit Bayesians, channeling their "Bayesian id" (Cohen, 1994). Bayes' theorem, formulated by Reverend Thomas Bayes in the 18th century, reminds us that in addition to the research evidence for a theory, the a priori likelihood of a theory should also count when evaluating it. Similarly, when evaluating whether a therapy possesses specific efficacy—do its ostensibly distinctive components really work?—theoretical plausibility matters. When the theoretical basis for a treatment is dubious or contradicts what we know, we should require more convincing research evidence for positive outcomes than if its theoretical basis is well established (Atwood, 2008; Goodman, 1999). This Bayesian way of thinking is exemplified by Hume's maxim, framed in popular terms by sociologist Marcello Truzzi (1978) and later, astronomer Carl Sagan (1995), as the mandate that "extraordinary claims require extraordinary evidence." The evidentiary bar should shift based on the initial theoretical plausibility.

Indeed, one problem with the current EST approach to ascertaining the evidence for psychotherapies—and I believe that David and Montgomery are on to something here—is that it adopts a classical "frequentist" probability approach to statistics, wherein ascertaining the evidence for a therapy is a relatively straightforward matter of "adding up" supportive evidence (either in terms of number of significant studies in box-score fashion, or summarizing effect sizes meta-analytically) rather than weighing such evidence in the context of a priori theoretical plausibility (see Ghaemi, 2009). From a Bayesian standpoint, this approach is missing something deeply important, especially if we care about specific efficacy, the efficacy of ostensibly distinctive therapeutic ingredients. A frequentist approach to probability leads to the conclusion, which many people understandably find problematic, that two supportive RCTs for thought field therapy should count as much evidence-wise as two supportive RCTs for a slightly new variant of cognitive-behavioral therapy. A Bayesian approach, in contrast, does not lead to this conclusion: it demands far more persuasive evidence to accept claims of efficacy, especially specific efficacy, when the theoretical plausibility of the

treatment (in this case, that a treatment removes blockages in invisible energy fields) is exceedingly low. Theoretical plausibility, in turn, may derive from many sources, including internal coherence of the theory, consistency of the theory with well-established scientific evidence ("connectivity"; Stanovich, 2009), and previous tests of the theory.

Although not explicitly invoking Reverend Bayes, my Ph.D. mentor David Lykken (1968) captured this concept in a somewhat different context, namely, psychological assessment. Lykken described an article by Sapolsky (1964) that was designed to test the idea that patients with certain psychiatric disorders, including eating disorders, harbor an unconscious "cloacal theory of birth," which posits that babies are born, much as a stool is passed, from a shared urinal, fecal, and reproductive orifice. According to Sapolsky, this notion generates the prediction that patients with eating disorders should more frequently report seeing cloacal animals, such as frogs, on the Rorschach Inkblot Test. Indeed, in a chart study, Sapolsky reported that significantly more patients with eating disorders than patients without eating disorders were Rorschach frog responders. When Lykken asked 20 of his academic psychology colleagues to estimate the likelihood of Sapolsky's cloacal theory being correct—before informing them of Sapolsky's findings—their estimates ranged from 10^{-6} to 0.13, with a median of 0.01 (which, as Lykken noted, we can translate roughly into "I don't believe it"). Yet, after Lykken informed his colleagues of Sapolsky's findings, they reestimated the plausibility of his theory as ranging between 10^{-5} and 0.14, with the median unchanged at 0.01 ("I still don't believe it"). From a purely frequentist standpoint, the psychologists may have been acting irrationally, as they were all but ignoring Sapolsky's positive results. From a Bayesian standpoint, they were not, as their initial estimates of the plausibility of Sapolsky's cloacal birth theory were so exceedingly low that a lone supportive study did virtually nothing to alter their subjective probabilities (see also Meehl & Rosen, 1955).

The same holds, I maintain, for evaluating the evidence for therapeutic efficacy. If a theory's a priori plausibility is exceedingly low, one should demand much more evidence for the specific efficacy of a therapy derived from it than if its plausibility is moderate

or high. As Sehon and Stanley (2003) argued in the context of medical treatments, “RCTs cannot stand apart from basic science. Even when a clinical trial returns positive results in the treatment arm that satisfy tests of statistical significance, we will have more confidence in these results when they have some antecedent biological plausibility” (p. 14). The same should hold for psychological plausibility. A Bayesian approach, I maintain, sidesteps many of the difficulties associated with explicitly incorporating research evidence for a theory into a classification scheme for evidence-based treatments. Among other things, it does not imply that a treatment with an absent or unsupported theoretical rationale should necessarily be relegated to a lower rung in the hierarchy of evidence-based treatments; it implies only that the bar for such a treatment be set higher than for treatments with a more plausible rationale.

Moreover, a Bayesian approach may be especially well suited for separating scientific from pseudoscientific treatments (see Lilienfeld, Lynn, & Lohr, 2003) because the theoretical plausibility of the latter tends to be low. Such an approach may not help us all that much in evaluating the evidence for “bona fide” therapies (Wampold et al., 1997), such as behavioral, cognitive-behavioral, and interpersonal therapies, as there remains much legitimate scientific debate about the plausibility of efficacious treatment mechanisms. But it should be helpful in pinpointing treatments that derive from an a priori unlikely theoretical rationale, and therefore require especially convincing research evidence before we can accept them into the corpus of evidence-based interventions.

Because of space constraints—and because doing so would require a separate manuscript of its own—I will not hazard an attempt at a formal classification scheme of evidence-based psychotherapies derived from Bayesian considerations (but see Goodman, 1999, for quantitative approaches). Nor will I discuss the potential limitations of a Bayesian approach, including the fact that the deductive linkages between theory and therapeutic technique are often modest. Suffice it to say that the Division 12 EST criteria and related classification schemes, although a valuable step in the direction of minimizing uncertainty in our clinical inferences, are insufficient. To separate the wheat from the chaff in

the psychotherapy field, we cannot evaluate treatment process or outcome research in a vacuum. As Reverend Bayes would have reminded us, we must consider such data in conjunction with the plausibility of our theoretical models.

REFERENCES

- Atwood, K. (2008). Prior probability: The dirty little secret of “evidence based alternative medicine.” *Science-Based Medicine*. Retrieved September 28, 2010, from <http://www.sciencebasedmedicine.org/?p=48>
- Beck, A. T. (1991). Cognitive therapy: A 30-year retrospective. *American Psychologist*, *46*, 368–375.
- Boisvert, C. M. (2010). Negative treatment effects: Is it time for a black-box warning? *American Psychologist*, *65*, 680–681.
- Borkovec, T. D. (1985). Placebo: Defining the unknown. In L. White, B. Tursky, & G. E. Schwartz (Eds.), *Placebo: Theory, research, and mechanisms* (pp. 59–64). New York: Guilford Press.
- Chambless, D. L., & Hollon, S. (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.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997–1003.
- David, D., & Montgomery, G. H. (2011). The scientific status of psychotherapies: A new evaluative framework for evidence-based psychosocial interventions. *Clinical Psychology: Science and Practice*, *18*, 89–98.
- Davidson, P., & Parker, K. (2001). Eye movement desensitization and reprocessing (EMDR): A meta-analysis. *Journal of Consulting and Clinical Psychology*, *69*, 305–316.
- Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist*, *65*, 21–33.
- Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., et al. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, *74*, 658–670.
- Eisner, D. A. (2000). *The death of psychotherapy: From Freud to alien abductions*. Westport, CT: Praeger.
- Feinstein, D. (2008). Energy psychology: A review of the preliminary evidence. *Psychotherapy: Theory, Research, Practice, Training*, *45*, 199–213.

- Ghaemi, S. N. (2009). *A clinician's guide to statistics and epidemiology in mental health: Measuring truth and uncertainty*. Cambridge, MA: Cambridge University Press.
- Goodman, S. N. (1999). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine*, *130*, 1005–1013. Retrieved September 30, 2010, from <http://www.annals.org/>
- Herbert, J., Lilienfeld, S. O., Lohr, J. M., Montgomery, R. W., O'Donohue, W., Rosen, G., et al. (2000). Science and pseudoscience in the development of eye movement desensitization and reprocessing: Implications for clinical psychology. *Clinical Psychology Review*, *20*, 945–971.
- Hunsley, J., & DiGuilio, G. (2002). Dodo bird, phoenix, or urban legend? *The Scientific Review of Mental Health Practice*, *1*, 11–22. Retrieved September 13, 2010, from <http://www.srmhp.org/>
- Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, *1*, 138–156.
- Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research*, *19*, 418–428.
- 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.
- Lilienfeld, S. O. (1998). Pseudoscience in contemporary clinical psychology: What it is and what we can do about it. *The Clinical Psychologist*, *51*(4), 3–9. Retrieved October 2, 2010, from <http://www.tandf.co.uk/journals/titles/13284207.asp>
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*, 53–70.
- Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2003). *Science and pseudoscience in clinical psychology*. New York: Guilford Press.
- Lohr, J. M., Lilienfeld, S. O., Tolin, D. F., & Herbert, J. D. (1999). Eye movement desensitization and reprocessing: An analysis of specific versus nonspecific treatment factors. *Journal of Anxiety Disorders*, *13*, 185–207.
- Lohr, J. M., Olatunji, B. O., Parker, L., & DeMaio, C. (2005). Experimental analysis of specific treatment factors: Efficacy and practice implications. *Journal of Clinical Psychology*, *61*, 819–834.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151–159.
- McCaslin, D. (2009). A review of efficacy claims in energy psychology. *Psychotherapy: Theory, Research, Practice, Training*, *49*, 249–256.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*, 194–216.
- Murphy, R., Cooper, Z., Hollon, S. D., & Fairburn, C. G. (2009). How do psychological treatments work? Investigating mediators of change. *Behaviour Research and Therapy*, *47*, 1–5.
- Ost, L. G., & Sterner, U. (1987). Applied tension. A specific behavioral method for treatment of blood phobia. *Behaviour Research and Therapy*, *25*, 25–29.
- Pignotti, M., & Thyer, B. (2009). Some comments on “Energy psychology: A review of the evidence”: Premature conclusions based on incomplete evidence? *Psychotherapy: Theory, Research, Practice, Training*, *49*, 257–261.
- Popper, K. (1959). *The logic of scientific discovery*. New York: Basic Books.
- Quine, W. V. O. (1951). Main trends in recent philosophy: Two dogmas of empiricism. *The Philosophical Review*, *60*, 20–43. Retrieved October 1, 2010, from <http://philreview.dukejournals.org/>
- Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification*, *27*, 300–312.
- Ruscio, J. (2006). *Critical thinking in psychology: Separating sense from nonsense* (2nd ed.). Pacific Grove, CA: Wadsworth.
- Sagan, C. (1995). *The demon-haunted world: Science as a candle in the dark*. New York: Random House.
- Sapolsky, A. (1964). An effort at studying Rorschach content symbolism: The frog response. *Journal of Consulting Psychology*, *28*, 469–472.
- Saravi, F. D. (1999). Energy and the brain: Facts and fantasies. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 43–58). New York: Wiley.
- Sehon, S. R., & Stanley, D. E. (2003). A philosophical analysis of the evidence-based medicine debate. *BMC Health Services Research*, *3*, 1–10.
- Shapiro, F. (1998). Eye movement desensitization and reprocessing (EMDR): Accelerated information processing and affect-driven constructions. *Crisis Intervention*, *4*, 145–157. Retrieved October 11, 2010, from <http://www.ingentaconnect.com/content/tandf/gcit>
- Shedler, J. (2010). The efficacy of psychodynamic therapy. *American Psychologist*, *65*, 98–109.

- Stanford, K. (2009). Underdetermination of theory. *Stanford Encyclopedia of Philosophy*. Retrieved October 11, 2010, from <http://plato.stanford.edu/entries/scientific-underdetermination/>
- Stanovich, K. (2009). *How to think straight about psychology* (9th ed.). Boston: Pearson.
- Taylor, S. (2007). Electroconvulsive therapy: A review of history, patient selection, technique, and medication management. *Southern Medical Journal*, *100*, 494–498.
- Truzzi, M. (1978). On the extraordinary: An attempt at clarification. *Zetetic Scholar*, *1*, 11–22. Retrieved September 9, 2010, from <http://www.tricksterbook.com/truzzi/Zetetic-Scholars.html>
- Tryon, W. W. (2005). Possible mechanisms for why desensitization and exposure therapy work. *Clinical Psychology Review*, *25*, 67–95.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Hillsdale, NJ: Erlbaum.
- 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.

Received October 18, 2010; accepted January 21, 2011.