Risky Tests of Etiological Models in Psychopathology Research: The Need for Meta-Methodology

Scott O. Lilienfeld & Melissa D. Pinto

Published online: 28 Aug 2015.

Submit your article to this journal

Article views: 99

View related articles

View Crossmark data

Risky Tests of Etiological Models in Psychopathology Research: The Need for Meta-Methodology

Scott O. Lilienfeld and Melissa D. Pinto

Department of Psychology, Emory University, Atlanta, Georgia

Progress in elucidating the causes of mental illness has been frustratingly slow (Insel, 2009). Hence, we welcomed Vaidyanathan, Vrieze, and Iacono's (this issue) provocative and well-reasoned essay calling for reform in the modal ways of doing business in psychopathology research. To be sure, some of this slow progress stems from inherent challenges in our subject matter (Meehl, 1978). For example, the causes of most or all major mental disorders appear to be exceedingly multifactorial (Kendler, 2005), and a full understanding of these causes necessitates a consideration of multiple levels of analysis ranging from the molecular to the physiological to the personological to the sociocultural (Lilienfeld, 2007). Needless to say, these difficulties are daunting. At the same time, as Vaidyanathan et al. note, the glacial progress of psychopathology research is also in large measure a self-inflicted wound. Perhaps in part for self-serving reasons, such as an understandable reluctance to subject our favored theories to intense scrutiny, we have often been reluctant to undertake "risky tests" of our models—those that place them at grave theoretical risk by maximizing the odds that they will be falsified if they are indeed false (Meehl, 1978; Popper, 1959).

We concur with all of Vaidyanathan et al.'s principal arguments and recommendations. For example, we share their well-stated concerns regarding "statisticism" (Duncan, 1984) and "technomyopia" (Tavris, 2012; see also Satel & Lilienfeld, 2013). Although we applaud the increasing migration of advanced statistical procedures, such as structural equation modeling and confirmatory factor analysis, into psychological and psychiatric journals, we worry that the growing popularity of techniques may at times engender the illusion of methodological rigor in its absence. In particular, these methods may unjustifiably reassure investigators that they can aggregate suboptimal indicators of psychopathology into latent variables, thereby circumventing shortcomings with their designs and measures. Hence, an overreliance on these methods may inadvertently generate a misleading sense of comfort with the research status quo and a further reluctance to undertake risky tests of theoretical models.

In some ways, this paradoxical result reminds us of the well-documented "risk compensation effect" in the health psychology literature (Pinkerton, 2001). Just as mandatory seat belt laws may sometimes backfire by reassuring automobile drivers that they are safe and can therefore drive dangerously (Evans & Graham, 1991), the use of high-level statistical techniques may sometimes backfire by reassuring researchers that they can always salvage less than adequate methodology by resorting to high-wire acts of impressive quantitative sophistication. But one cannot make a silk's purse out of a sow's ears. Advanced statistical methods, as valuable as they are, cannot substitute for carefully conceptualized research designs, well-validated measures, and clinically significant translational research that generalizes to real-world settings outside the laboratory, nor can they allow for risky tests of theories in the absence of internally valid designs, those that rule out many or most rival threats to inferring causality.

Indeed, we enthusiastically second Vaidyanathan et al.'s exhortation to use research designs, especially genetically informed designs, which permit risky tests of causal hypotheses. If we hold a different perspective from Vaidyanathan et al., it is almost certainly one of degree rather than of kind. Specifically, we suspect that we place somewhat more weight than they do on *meta-methodological* issues (Turner & Durham, 2014)—those that bear on the evaluation of theories across multiple studies—in the appraisal of etiological hypotheses. In fairness, Vaidyanathan et al. offer a brief nod to issues of replicability, multiple testing, and research practices that can lead to false positive results (p. 5), and it is evident that they are well aware of the importance of meta-methodological issues in theory appraisal.

Because of the serious inferential problems posed by auxiliary hypotheses (e.g., the measures used, the samples examined; see Meehl, 1978) in the "soft" areas of psychology, such as psychopathology research, however, there are precious few, if any, definitive studies or "experimenta crucis" (Lohne, 1968): isolated investigations that definitely refute a theory. Instead, refutation of theories in psychopathology requires the gradual accumulation of negative evidence across multiple well-conducted studies, just as corroboration of theories requires an accumulation of supportive evidence across many studies.

In this brief commentary, we therefore offer a friendly amendment to Vaidyanathan et al.'s superb commentary by touching on three meta-methodological considerations that we regard as crucial to subjecting etiological models of psychopathology to risky tests (see also Lilienfeld, 2004): replicability, use of converging indicators, and specificity designs. Of course, none of these meta-methodological consideration is new, but we contend that they have continued to receive inadequate attention in many discussions of psychopathology research methodology. We also highlight a feature of risky tests that Vaidyanathan et al. do not emphasize explicitly but that we believe complements their central thesis. Specifically, we underscore the point that risky tests, by helping to eliminate rival hypotheses, minimize the chances that investigators will be fooled into accepting spurious conclusions regarding the etiology of mental disorders.

Replicability

Much has been written about the vital importance of replicability over the past decade (e.g., Ioannidis, Asendorph et al., 2005, 2013; Nosek, Spies, & Motyl, 2013), and we see no need to retread that well-worn ground here. At the same time, we believe that the value of replication continues to receive short shrift in discussions of risky hypothesis tests in psychopathology research. One recent survey revealed that the approximate replication rate for findings in psychology appears to be about 1% (Makel, Plucker, & Hegarty, 2012). Given the difficulty of recruiting participants with serious psychological disorders, this percentage may actually be overly sanguine for psychopathology research. Moreover, funding agencies have traditionally accorded higher priority to the "sexiness" of potential findings than to replication of previous ones (e.g., Hartschorne & Schachner, 2012). Without a culture of replicability in psychopathology research, theories will not be subjected to grave theoretical risk, because false-positive results that appear to support a theory will be allowed to persist in the literature without self-correction.

In an article that should be required reading for all psychopathology students and researchers, Lykken (1968) distinguished among three forms of replication: literal, operational, and constructive. In literal replications, one more or less replicates precisely the sampling and measurement techniques of the original investigator; as Lykken noted, this form of replication is rarely possible, although running additional participants from the initial sample is a close approximation. In operational replications, one uses the same measurement operations as the original investigator, in essence adopting the investigator's "recipe" for obtaining positive results as a guide. In constructive replications, one examines the robustness of the effects by extending the initial investigators' findings to alternative measures of the same construct, hence the term "constructive." Constructive replications, now more commonly termed conceptual replications (Hendrick, 1990), are critical in psychological science, as they help to ascertain whether a theory's predictions hold across differing measures and experimental conditions.

A lively debate has recently spilled over to the pages of psychological journals concerning the importance of direct replication, which is probably closest to Lykken's (Lishner, 1968) concept of operational replication. Some authors (e.g., 2015; Simons, 2014) believe that direct replications of findings, in which investigators determine whether a researchers' methodological recipe yields positive results, are essential to a mature psychological science, whereas others (e.g., Stroebe & Strack, 2014) believe that the genuine crucible of a theory's mettle is conceptual replication and that direct replications are frequently unnecessary. We come down squarely on the side of the former camp. As we discuss in the following section (Use of Converging Indicators), conceptual replications are extremely valuable in examining the generalizability of a theory's predictions to diverse indicators. Nevertheless, without first establishing the robustness of an initial effect, conceptual replications can be misleading. One must first ascertain whether an effect is genuine before ascertaining its boundary conditions.

In particular, a focus on conceptual replication in the absence of direct replication can produce what we term "the illusion of replicability." To generate this illusion, investigators testing a theory may first obtain positive results for Variable X. Presuming that this finding is replicable, they then examine allied Variable Y and again obtain positive results. The process repeats itself in a third study for Variable Z. Nevertheless, had the investigators examined Variables X, Y, and Z across all three studies, they might well have obtained negative results for two of the three variables. For example, many advocates of the storied Rorschach Inkblot Test, a measure that continues to be widely used in studies of the correlates and etiology of psychopathology, have contended that it is well suited for the detection of suicide risk (Mihura, Meyer, Dumitrascu, & Bombel, 2013). Nevertheless, a close inspection of studies on the Rorschach and suicide risk reveals that this conclusion rests largely on four studies—two on completed suicides, one on individuals who attempted suicide, and one on a serotonin metabolite has been linked to suicide risk (Wood, Garb, Nezworski, Lilienfeld, & Duke, 2015). Nevertheless, because one of the studies on completed suicides (Exner & Wylie, 1997) was

essentially a "fishing expedition" that almost surely capitalized on chance, the genuine body of data most likely rests on three studies, each using a different index of suicide risk. Before one can legitimately conclude that the Rorschach detects risk for suicide, multiple direct replications using the same dependent measures will be necessary (see also Chaplin & Goldberg, 1984, for an example of the illusion of replicability in the literature on moderators of the cross-situational consistency of behavior).

The bottom line is that conceptual replication in the absence of direct replication can, paradoxically, subject a theory to *lower* theoretical risk, as it can allow investigators who are partial to a theory to claim that this theory has been corroborated when such corroboration is spurious. Hence, multiple direct replications of findings by independent investigative teams are essential for risky tests in psychopathology research.

Use of Converging Indicators

Closely allied to the issue of conceptual replication is the use of converging measurement indicators on both the outcome and predictor ends of our equations (Stanovich, 2012). Most psychopathology researchers appear to be aware that findings relevant to etiological models should ideally hold across varying measures and experimental conditions. Nevertheless, there has been a troubling tendency for researchers to refer to certain well-validated measures of psychopathology, such as the Psychopathy Checklist–Revised (Ermer, Kahn, Salovey, & Kiehl, 2012) or the Hamilton Depression Rating Scale (Bagby, Ryder, Schuller, & Marshall, 2014), as "gold standards."

As Cronbach and Meehl (1955) pointed out 60 years ago, there are few, if any, genuine gold standards in psychopathology research (see also Faraone & Tsuang, 1994; Skeem & Cooke, 2010), because all measures of psychopathology are necessarily fallible indicators of the constructs of interest. The exceedingly rare exceptions to this rule that we might carve out are for conditions initially believed to be "mental disorders" for which an organic pathology and etiology was later discovered, such as general paresis (tertiary syphilis) and pellagra, the former by caused the syphilis spirochete and latter by niacin deficiency (Kendler, 2006). For these few conditions, physicians have developed laboratory tests that we can regard as approaching gold standards. In contrast, no contemporary psychological disorders can be diagnosed using laboratory assays and must instead be assessed by a process of triangulation across multiple converging indicators.

The confusion here is more than terminological, because in some research laboratories certain well validated measures, such as the PCL-R, the Beck

Depression Inventory, or Structured Clinical Interview for DSM Disorders, have become accepted as "the" standard benchmarks of the psychopathological conditions of interest. As a consequence, investigators in these laboratories forfeit the opportunity to subject their preferred etiological models to more stringent tests by examining the boundary conditions of these models. Even worse, these researchers may at times arrive at inaccurate conclusions, because their positive findings may be a spurious consequence of an idiosyncrasy of the specific measure used and may not generalize to other measures. For example, in research on psychopathic personality (psychopathy), one is left to wonder how many of the findings bearing on prominent etiological models are relevant primarily to nonspecific antisocial or criminal behaviors rather than the core interpersonal and affective traits of psychopathy per se (e.g., guiltlessness, callousness, fearlessness) given that most widely used measures of this condition are heavily saturated with these behaviors (Skeem & Cooke, 2010).

The use of multiple indicators of both disorders and dependent measures allows for *multiple corroboration* of a theory's predictions across studies (Lykken, 1968), which in our view is among the most critical means of subjecting etiological theories to risky tests. Cook's (2000) often overlooked principle of the "heterogeneity of irrelevancies" is relevant in this context. No psychological measure is free of error. Nevertheless, to the extent that a theory of etiology generates predictions that are corroborated across measures with largely offsetting errors, our confidence in this theory is justifiably buttressed. The principle of the heterogeneity of irrelevancies provides a potent rationale for administering measures using markedly different modes of assessment (e.g., self-report, interview, informant report, psychophysiology), as measures derived from different modes are likely to possess largely independent errors. Industrial/organizational psychologists have long embraced the importance of validity generalization (Schmidt & Hunter, 1977), but this concept appears to have received insufficient application to evaluating the robustness of hypotheses in psychopathology research.

The failure of an etiological theory to generalize across operationalizations of the disorder or experimental conditions does not necessarily imply that the theory is false. It may indicate that the theory is correct but only within certain boundary conditions, or it may indicate that the ostensibly similar measures of the same construct are actually assessing different constructs (e.g., Kagan, 1988), an error known as the jangle fallacy (Block, 1995).

Although we share Vaidyanathan et al.'s enthusiasm for latent variable models to capitalize on the "signal" within psychopathology indicators that cuts

across the "noise" induced by measurement error, we contend that these models should supplement, not substitute for, careful analyses of the validity generalization of findings. We worry that many researchers too quickly move to create latent variables out of varying indicators of a construct without first examining whether the observed correlations are reasonably consistent across these indicators. If they are not, this finding may again point to shortcomings of the model or the need to identify boundary conditions for its applicability.

One commendable example of the use of converging indicators on the dependent variable side of the equation derives from research on cognitive biases in anxiety disorders. Research using a variety of paradigms that present participants with threat cues, such as homophone disambiguation tasks, emotional Stroop tasks, visual looming tasks, and dot-probe tasks, reveals that individuals with anxiety disorders tend to be overly attentive to signals of threat. Moreover, these findings have often have been replicated for each of these tasks (McNally, 1996). At the same time, there appear to be important boundary conditions on this association, with the relation between anxiety and attentional hypersensitivity holding for certain laboratory tasks but not others, and more for generalized anxiety disorder than for high levels of "normal-range anxiety" (van Bockstaele et al., 2014).

Specificity Designs

Not uncommonly, psychopathological researchers who intend to test an etiological model will proceed in the following fashion. They begin by identifying a group of individuals with Disorder X. They then identify a group of matched ostensible normal "controls" (who should really be termed "normal comparison participants," as they are not strict controls), often matching them on various ostensible nuisance variables (e.g., social class, measured intelligence, gender). Finally, they compare these two groups on a presumed etiological risk factor. If the groups differ significantly, the authors conclude that the ostensible variable is now supported as a potential etiological risk factor for Disorder X. Indeed, many articles in abnormal psychology and psychiatry journals continue to compare groups of individuals with a given disorder, such as borderline personality disorder, with normal comparison participants (e.g. Herpertz, Kunert, Schwenger, & Sass, 2014).

Aside from the well-known inferential and statistical problems associated with matching and covariate control procedures (see Meehl, 1971; Miller & Chapman, 2001), design is becoming increasingly difficult

to defend as a stand-alone method, because it rarely subjects etiological theories of psychopathology to more than minimal risk. Specifically, it does not permit researchers to ascertain whether a putative risk factor is unique to the condition in question as opposed to (a) psychopathology in general or (b) one or more specific disorders.

To begin to address this question and subject etiological theories to greater risk, specificity designs are required. These designs compare individuals with Disorder X either with (a) a group of mixed psychiatric patients or (b) one or more disorders that overlap phenotypically with Disorder X but that are posited not to share the etiological risk factor in question. In general, design (a) is better suited to ascertaining whether a variable is specific to Disorder X than to psychopathology in general, whereas design (b) is better suited to ascertaining whether a variable is specific to Disorder X as opposed to Disorders Y or Z, which the investigator believes are different in etiology from Disorder X (Garber & Hollon, 1991). We regard the need for specificity designs as a meta-methodological consideration because rarely, if ever, can an investigator satisfactorily address questions of disorder etiology with a single, "crucial" study. Instead, only by comparing the disorder of interest with differing psychiatric comparison groups across studies can an investigator gradually work toward establishing that variable as a marker of *specific etiology* (Meehl, 1977) for that disorder. In this respect, the oft-asked question of "Which is the right comparison group?" is almost always ill-conceived, because different comparison groups address different questions.[1]

Recent factor-analytic research on the potential "general factor" (p factor) of psychopathology (Lilienfeld, Caspi et al., 2014) underscores the need for specificity designs, because it highlights the point that the covariation among different mental disorders (often called "comorbidity," a term that is typically presumptuous because it assumes that the conditions in question have been established as pathologically and etiologically separable; Waldman, & Israel, 1994) or among continuously measured features of these disorders is far more often the rule than the exception. As a consequence, discriminant validity is typically at least as important as convergent validity when undertaking risky tests of etiological models of psychopathology (Lilienfeld, 2004). In reality, the general factor of psychopathology may be nothing

---

[1]One of the few cases in which the comparison of a psychiatric disorder with a group of ostensible normals is warranted is when a theoretical model posits that the etiological variable is not related specifically to the disorder in question but to psychopathology in general. Even then, however, the investigator will also need to examine psychiatric comparison samples to ensure that the variable cuts across most or all forms of mental disorder.

new; it is well-known that the higher order dimension of negative emotionality pervades virtually all forms of serious psychopathology, with only a few potential exceptions (e.g., psychopathy, bipolar disorder; Tellegen & Waller, 2008; Watson & Clark, 1984). Lykken (1991) similarly referred to the "crud factor," the modest ambient level of psychological covariation (roughly analogous to the cosmic microwave background) stemming from the loose tendency for most measures of individual differences, including those in the psychopathology domain, to hang together. One primary contributor to the crud factor in psychopathology research is almost surely the propensity for virtually all measures of mental disorder to be saturated with negative emotionality. For these reasons, specificity designs are crucial for subjecting etiological hypotheses regarding psychopathology to theoretical risk. Without them, investigators can be deceived into believing that their hypotheses concerning specific etiology have been corroborated when they have not.

The finding that a variable is not specific to a given disorder but is instead relevant to multiple disorders does not necessarily imply that this variable is etiologically unrelated to the disorder (Garber & Hollon, 1991). For example, life stressors appear to boost risk for numerous conditions, including schizophrenia and depression (Walker & DiForio, 1997), and child maltreatment is tied to heightened risk for numerous mental disorders (Caspi et al., 2014). Furthermore, compelling evidence ties chronic physical illnesses, especially those affecting the immune system, to mental illness, particularly depression, suggesting a potential shared etiological mechanism linking alterations in mood with chronic medical illnesses (Evans et al., 2005). Nevertheless, the finding that a putative risk variable is linked to numerous disorders may lead scholars to reconsider their etiological models, and in particular to reevaluate this variable's status as a marker of specific etiology (Meehl, 1977).

Concluding Thoughts

In conclusion, we heartily endorse Vaidyanathan et al.'s well-taken recommendations for subjecting etiological models of psychopathology to greater theoretical risk. By excluding plausible alternative hypotheses for findings, risky tests minimize the odds that investigators will be fooled into accepting etiological models that are mistaken or incomplete. In this commentary, we have argued that genuinely risky tests of etiological theories require assiduous attention to widely known but widely overlooked metamethodological desiderata, which apply not merely to individual studies but all to phases of an investigator's research program.

Note

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

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108–119.

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2014). The Hamilton depression rating scale: Has the gold standard become a lead weight? *American Journal of Psychiatry*, *161*, 2163–2177.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187–215.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., & Moffitt, T. E. (2014). The p factor one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*, 119–137.

Chaplin, W. F., & Goldberg, L. R. (1984). A failure to replicate the Bem and Allen study of individual differences in cross-situational consistency. *Journal of Personality and Social Psychology*, *47*, 1074–1090.

Cook, T. D. (2000). Toward a practical theory of external validity. In L. Bickman (Ed.), *Validity and social experimentation*: *Donald Campbell's legacy* (Vol. 1, pp. 3–43). Thousand Oaks, CA: Sage.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Duncan, O. D. (1984). *Notes on social measurement, historical and critical*. New York, NY: Russell Sage Foundation.

Ermer, E., Kahn, R. E., Salovey, P., & Kiehl, K. A. (2012). Emotional intelligence in incarcerated men with psychopathic traits. *Journal of Personality and Social Psychology*, *103*, 194–204.

Evans, D. L., Charney, D. S., Lewis, L., Golden, R. N., Gorman, J. M., Krishnan, K. R., & Valvo, W. J. (2005). Mood disorders in the medically ill: Scientific review and recommendations. *Biological Psychiatry*, *58*, 175–189.

Evans, W. N., & Graham, J. D. (1991). Risk reduction or risk compensation? The case of mandatory safety-belt use laws. *Journal of Risk and Uncertainty*, *4*, 61–73.

Faraone, S., & Tsuang, M. (1994). Measuring diagnostic accuracy in the absence of a "gold standard." *American Journal of Psychiatry*, *151*, 650–657.

Garber, J., & Hollon, S. D. (1991). What can specificity designs say about causality in psychopathology research? *Psychological Bulletin*, *110*, 129–136.

Hartshorne, J. K., & Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience*, *6*, 8.

Hendrick, C. (1990). Replications, strict replications, and conceptual replications: are they important? *Journal of Social Behavior & Personality*, *5*, 41–49.

Herpertz, S. C., Kunert, H. J., Schwenger, U. B., & Sass, H. (2014). Affective responsiveness in borderline personality disorder: A psychophysiological approach. *American Journal of Psychiatry*, *156*, 1550–1556.

Insel, T. R. (2009). Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. *Archives of General Psychiatry*, *66*, 128–133.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Kagan, J. (1988). The meanings of personality predicates. *American Psychologist*, *43*, 614–620.

Kendler, K. S. (2005). Toward a philosophical structure for psychiatry. *American Journal of Psychiatry*, *162*, 433–440.

Kendler, K. S. (2006). Reflections on the relationship between psychiatric genetics and psychiatric nosology. *American Journal of Psychiatry*, *163*, 1138–1146.

Lilienfeld, S. O. (2004). Taking theoretical risks in a world of directional predictions. *Applied and Preventive Psychology*, *11*, 47–51.

Lilienfeld, S. O. (2007). Cognitive neuroscience and depression: Legitimate versus illegitimate reductionism and five challenges. *Cognitive Therapy and Research*, *31*, 263–272.

Lilienfeld, S. O., Waldman, I. D., & Israel, A. C. (1994). A critical examination of the use of the term and concept of comorbidity in psychopathology research. *Clinical Psychology: Science and Practice*, *1*, 71–83.

Lishner, D. A. (2015). A concise set of core recommendations to improve the dependability of psychological research. *Review of General Psychology*, *19*, 52–68.

Lohne, J. A. (1968). Experimentum crucis. *Notes and Records of the Royal Society of London*, *23*, 169–199.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151–159.

Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Cicchetti, & W. M. Grove (Eds.), *Thinking clearly about psychology Volume 1: Matters of public interest* (pp. 3–39). Minneapolis: University of Minnesota Press

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*, 537–542.

McNally, R. J. (1996). Cognitive bias in the anxiety disorders. In *Nebraska symposium on motivation* (Vol. 43, pp. 211–250). Lincoln, NE: University of Nebraska Press.

Meehl, P. E. (1971). High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology*, *77*, 143–148.

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.

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

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*, 40–46.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615–631.

Pinkerton, S. D. (2001). Sexual risk compensation and HIV/STD transmission: Empirical evidence and theoretical considerations. *Risk Analysis*, *21*, 727–736.

Popper, K. (1959). *The logic of scientific discovery*. London, UK: Hutchinson.

Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed*: *The seductive appeal of mindless neuroscience*. New York, NY: Basic Books.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*, 529–540.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80.

Skeem, J. L., & Cooke, D. J. (2010). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment*, *22*, 433–445.

Stanovich, K. (2012). *How to think straight about psychology* (10th ed.). Boston, MA: Pearson.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*, 59–71.

Tavris, C. A. (2012, May). *Debunking pseudoneuroscience*. APS David Myers Distinguished Lecture on the Science and Craft of Teaching Psychology, annual convention of the Association for Psychological Science, Chicago, IL.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. *The SAGE Handbook of Personality Theory and Assessment*, *2*, 261–292.

Turner, J. R., & Durham, T. A. (2014). Meta-methodology: Conducting and reporting meta-analyses. *The Journal of Clinical Hypertension*, *16*, 91–93.

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological Bulletin*, *140*, 682–721.

Walker, E. F., & Diforio, D. (1997). Schizophrenia: A neural diathesis-stress model. *Psychological Review*, *104*, 667–685.

Watson, D., & Clark, L. A. (1984). Negative affectivity: the disposition to experience aversive emotional states. *Psychological Bulletin*, *96*, 465–490.

Wood, J. M., Garb, H. N., Nezworski, M. T., Lilienfeld, S. O., & Duke, M. C. (2015). A second look at the validity of widely used Rorschach indices: Comment on Mihura, Meyer, Dumitrascu, and Bombel (2013). *Psychological Bulletin*, *141*, 236–249.