aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Do You Believe the Field of Psychological Science Is Headed in the Right Direction?

Perspectives on Psychological Science
2017, Vol. 12(4) 675–679

DOI: 10.1177/1745691617706507
www.psychologicalscience.org/PPS
SAGE

Robert W. Levenson
Department of Psychology, University of California, Berkeley

Abstract
The state of psychological science is considered in terms of current issues and suggestions for the future.

Keywords
replication, research design, careers

Viewed from afar, psychology appears to be adrift in troubled seas. Although world-weary cynics might say "it's always something," the current state of affairs in psychological science is anything but ordinary. One dominant feature of this seascape is the "replicability crisis," which has raised disturbing questions about the reproducibility of the field's foundational science. A number of thoughtful analyses (e.g., Open Science Collaboration, 2015) have provided powerful evidence that something is truly amiss and point fingers of culpability in numerous directions, including researchers (for carelessness, scientific myopia, filling file drawers with null findings, and overvaluing *p* levels); the profession (for unrelenting publication pressures in hiring and promotion); and journals (for biases against publishing null results and direct replications, and, in some cases, for preferring sizzle over substance). At a more macro level, we are in a period of declining federal funding for the most basic of behavioral science research. This has caused many of our most gifted psychological scientists to abandon the research problems they care about most deeply and, in some cases, to leave research completely. In areas of psychological science where funding has been preserved, there are growing discrepancies between funding "haves" and "have nots," with available funds being increasingly funneled into the coffers of more senior investigators (Daniels, 2015). This has created a difficult and often demoralizing situation for younger investigators, who are ready to take on important scientific problems with energy and cutting-edge methods but aren't senior enough to draw the funding they need. Moreover, with a new administration coming into power in the United States, there is uncertainty about the fate of science funding. And one final bit of gloom and doom: Tenured "hard money" positions in psychology, the safe havens that in theory afford scientists protection from the vicissitudes of funding and fate and free them up to conduct thoughtful and careful science, are becoming an increasingly endangered species. Withering economic realities confront colleges, universities, and medical centers, thus forcing them to rely increasingly on nontenured faculty. Moreover, all faculty, both tenured and nontenured, are viewed as potential profit centers, thrusting them into fund raising and entrepreneurial roles that are not always good fits for their abilities. Beneath all of these worrisome realities is the very real specter of psychological scientists being pushed and pulled toward conducting more expedient, strategic, conservative, careerist science and away from doing the most challenging, innovative, important, and courageous science possible.

The Replication Crisis: An Opportunity for Midcourse Correction

The replication crisis has spawned serious self-examination and a number of proposals (e.g., Nosek, Spies, & Motyl, 2012) for altering the ways that psychological science is conducted both at the macro level (e.g., open science initiatives, research registries, new journal policies) and at the level of the individual scientist (e.g., recommended best practices for data collection, analysis, and disclosure). These are clearly positive developments, and we owe a debt of gratitude to those who have taken on this difficult and, I expect, sometimes thankless task. Of course, with

Corresponding Author:
Robert W. Levenson, Department of Psychology, University of California, Berkeley, Tolman Hall #1650, Berkeley, CA 94720-1650
E-mail: boblev@berkeley.edu

the advantage of hindsight, the replication crisis was predictable. Early warnings about the importance of effect sizes and statistical power (Cohen, 1977), rigorous hypothesis testing (Platt, 1964), and unreported replication failures (Rosenthal, 1979) began appearing 40–50 years ago. In the ensuing decades, these ticking bombs led to a number of scattered explosions and even some changes in journal publication policies, but nothing that would qualify as a true sea change. Fast forward to the present, and these issues are being discussed in ways that are more vocal, public, widespread, organized, and action-oriented than they were in earlier discussions. Hopefully, we will look back at this period as a true inflection point—a moment when psychological science took a significant turn in the right direction.

Replicability: Necessary but Not Sufficient

If we could only clean up the replication problem, then all would be well: tempting but, of course, far from true. Improving the reproducibility of psychological science is clearly important; reproducibility, however, does not ensure quality. A study can comply with all of the proscriptions of open science (e.g., preregistered hypotheses, sample sizes that ensure appropriate power, methods and data publicly available, no cherry picking among variables and findings) and be supported by direct replication without shrinking effect sizes but still be trivial and unimportant. Much as reliability limits validity but does not ensure it, replicability reassures us that the science was done properly and that the findings are likely to be "real," but it does not guarantee that these findings will create new insights, add meaningfully to existing bodies of knowledge, provide critical tests among competing theories, generalize to other situations and populations, or address human suffering. Replicability is foundational and necessary but clearly is not sufficient for taking the full measure of psychological science.

Five Suggestions

With the current spate of self-examination and the growing mandate for change stimulated by the replication crisis, it is a good time to broaden the discussion beyond replicability per se to take stock of other opportunities for moving psychological science in the right direction. In the spirit of this special section, I offer five suggestions that might help accomplish this goal.

1. Take sampling more seriously

Psychological science is awash in samples that are constructed on the basis of "convenience" rather than being constructed to represent populations of interest. Nonrepresentative samples create enormous problems for external validity by sorely limiting our ability to generalize findings to larger populations. In addition, these samples can contribute to problems with replicability (e.g., when original and replication samples are not comparable). Psychological scientists rarely take the level of care in sample construction that is expected in other social sciences. Rather, a great deal of psychological science is done with the most convenient samples of all: undergraduates fulfilling psychology course requirements and online "workers." Understandably, the geographical limitations of laboratory research makes it impractical to construct representative national samples (although in recent years satellite laboratory studies have been grafted on to large representative national surveys), but samples can be constructed to be representative of local populations. Large bodies of research indicate that psychological processes are profoundly influenced by factors such as sex, ethnicity, age, socioeconomic status, education, and religion. Thus, it seems critical to make sure that these kinds of factors are taken into account when constructing samples (e.g., Graham, 1992). Careful sample construction can be difficult, time consuming, and expensive, and psychologists typically receive little training in the methods involved. In an era of widespread interest in psychological diversity, it will be increasingly important to know exactly who we are studying in our research and increasingly perilous to assume that different samples recruited by different laboratories are ipso facto equivalent on the basis of the most superficial of similarities (e.g., "college students" or "young adults").

2. Bite the bullet on longitudinal designs

Prediction is the centerpiece of psychological science. A primary goal of our research is to anticipate how lives will turn out. Although some things happen in an instant, many important psychological outcomes (health, well-being, life satisfaction, relationship satisfaction, competence, expertise, attitudes, values, psychopathology, etc.) develop and change over time. Most of the foundational questions in psychological science would benefit greatly from longitudinal designs (a point that is made in countless discussion sections when data are not available to disambiguate key directional and temporal relationships among variables). Studies in which predictors and outcomes are measured at the same point in time may be efficient and convenient and may help feed the hungry beasts of publication and career pressures. However, they ultimately leave us with intriguing findings that beg for additional information about directionality (e.g., if cancer progression shows a significant negative correlation with positive emotion, is it because positive emotion slows the progression of cancer and/or because having

cancer that is slowly progressing makes us feel better?). In many areas of psychological science, mediational analyses (e.g., Baron & Kenny, 1986) are considered de rigueur for understanding processes and mechanisms. To realize the full value of these analyses requires longitudinal designs in which predictors, mediators, and outcomes are measured in the appropriate temporal order. Finally, we have an embarrassment of methodological riches available for analyzing longitudinal data (e.g., Duncan, Duncan, Strycker, Li, & Alpert, 1999; Ferrer & Song, 2012; McArdle, 1988) with many new computational approaches and statistical tools that were not readily available in earlier times. Thus, there are enormous opportunities for leveraging longitudinal data to address some of the most important and enduring questions in psychological science.

3. Embrace clinical phenomena

In many health-related fields (e.g., neurology, psychiatry), research scientists receive applied clinical training as a matter of course. In psychology, clinical training (i.e., learning about the diagnosis and treatment of mental illness) is generally reserved for students in clinical psychology. Moreover, because clinical psychology is usually the most competitive area of psychology for graduate admissions, programs often develop policies to protect against "Trojan Horse" applicants who apply to other areas of psychology with the intention of switching to the clinical area later on. This protectionism creates resentment, intellectual silos, and proprietary fiefdoms. Add to this the commonly held view that clinical training takes valuable time away from students' research, and the end result is that students in other areas of psychology gain little exposure and experience working with clinical problems and populations. I consider this state of affairs to be extremely unfortunate and short-sighted (Levenson, 2014). Reducing the burden of mental illness is one of the primary "deliverables" for all of psychological science (e.g., peruse the significance sections of many grant proposals). Although the National Institute of Mental Health (NIMH) is often characterized as being unsupportive of basic behavioral science, two of its major statements of research priorities—"translational research" (National Institute of Mental Health, 1999) and the Research Domain Criteria (Cuthbert & Insel, 2010)—both envision mental illness as grounded in dysfunction in foundational psychological processes such as social connection, cognition, personality, emotion, motivation, development, and learning. In psychological science, the primary expertise and deepest knowledge regarding these processes typically resides in areas outside of clinical psychology.

In recent years, progress in preventing and curing mental illness has lagged far behind similar progress with physical illness (Insel & Gogtay, 2014; Levenson, in press). Simply stated, we have a number of effective ways to reduce the suffering associated with mental illness, but preventative interventions and cures have remained elusive. As a result, prevalence rates for mental illness remain stubbornly stuck (e.g., Kessler, Chiu, Demler, & Walters, 2005). Imagine for a moment if psychological scientists of all stripes devoted a least some of their scientific horsepower to addressing mental health problems. And imagine if graduate students, postdocs, and faculty from all areas of psychological science were welcomed when they expressed interest in working on clinical issues. Who knows how much progress could have been made toward reducing the crippling burden of mental illness, which is now among the most burdensome of all illnesses worldwide (Ferrari et al., 2013)?

4. Use stronger research designs

In thinking about improving the quality of future psychological science, it is worth remembering three "oldies but goodies": (a) scientific progress is based on theories and hypotheses that can be disconfirmed (Popper, 1959), (b) understanding in science progresses by posing and testing alternative hypotheses (Platt, 1964), and (c) testing rival scientific theories is not an "either–or" endeavor but rather is about establishing the boundary conditions under which theories do and do not hold (Jost & McGuire, 2013). One clear implication of these principles is that psychological science advances most rapidly under research conditions that enable our ideas to be proven wrong.

As a clinical psychologist by training, I am struck by the need to use stronger research designs in clinical research. Studies of psychopathology often compare individuals with a particular diagnosis with healthy controls, which is not optimal for revealing the features that distinguish one disorder from another (the latter benefits from research designs that include individuals with different forms of psychopathology). Treatment studies often compare a complex, multifaceted, manualized treatment with a "no-treatment" condition, typically finding that "something" does better than "nothing" in reducing symptoms of dysfunction. But this kind of design cannot tell us what parts of the active treatment were necessary or sufficient for producing the positive outcomes and which treatments (or parts of treatments) were most effective for treating different disorders. Fifty years ago, "dismantling" studies (which enable examining the impact of individual elements in a complex treatment package) were being conducted (Paul & Shannon, 1966), but these kind of research designs have unfortunately remained the exception rather than the rule. Recently, the NIMH weighed in on this issue by mandating that treatment studies utilize stronger research designs in which there is a designated target mechanism, a way of establishing that the target is

engaged by the treatment, and a determination of whether the target engagement mediates the therapeutic outcome (Insel, 2015). Clinical research is extraordinarily demanding and difficult. These challenges notwithstanding, this is surely an area of psychological science where using the strongest research designs possible is paramount.

5. Use stronger categorical variables

Although contemporary approaches often eschew group designs analyzed using ANOVAs in favor of continuous variables analyzed using regression, structural equations models, and the like, categorical variables still play an important role in psychological science. As with so many aspects of our science, convenience often works against quality. For example, in cultural psychology, nationality or self-designated ethnicity is often used as a convenient way to form groups; in clinical psychology, diagnoses from the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) are used. Using these categorical variables, a cross-cultural study might compare thoughts, feelings, or behaviors among African Americans, Chinese Americans, and European Americans, and a clinical study might do the same for different diagnostic groups. In both cases, problems abound. In the cross-cultural studies, individuals who self-identify with a particular ethnicity can differ greatly in their exposure to the cultural traditions of interest. This results in groups that are highly heterogeneous. Approaches that address these concerns require laborious consideration of cultural exposure and practices (e.g., inclusion and exclusion factors that consider birthplace of participants, parents, and grandparents; ethnic makeup of childhood and current social networks; and, as appropriate, cultural factors related to language, diet, and religion). Absent this kind of care in forming cultural groups, within-group heterogeneity can work against finding reliable group differences.

In clinical research, the use of *DSM* diagnoses continues to dominate published research despite a litany of criticism for within-diagnosis heterogeneity; arbitrary cut-offs for inclusion/exclusion; cross-diagnosis comorbidity of symptoms; and lack of utility for predicting etiology, course, or response to treatment. Group comparisons based on *DSM* diagnoses are at the core of almost all research on mental illness, including studies of psychological and pharmacological treatments, family and molecular genetics, neural circuitry, etiology, developmental course, and disease biomarkers. Time will tell whether new approaches that eschew *DSM* diagnoses (e.g., NIMH's Research Domain Criteria, which is based on specific symptoms, dimensions of functioning, and associated neural circuits; Insel et al., 2010) prove more productive in increasing knowledge and reducing the burden of mental illness.

The little five

These five suggestions represent a subset of issues where psychological science could benefit from serious discussion about best research practices and standards to help move our field in the right direction in the future. The replication crisis and its aftermath provide a good model for how these discussions can proceed in ways that are largely constructive and inclusive, and that can lead to positive changes in the ways that psychological science is conducted.

Advice for Younger Colleagues

A number of years ago, my Berkeley colleagues were discussing graduate training and mentoring in a faculty meeting. At one point, I asked for a show of hands as to whether people considered themselves "self-made" or "mentored." The results were striking; almost all of the hands were raised for "self-made." Given the uncertainties of the future, there is no way to know whether the practices, decisions, and strategies that led to success in the past will lead to success in the future. In addition, there are huge differences among individual scientists in their strengths and weaknesses, the contexts they work in, the compromises and sacrifices they are willing to make, and their standards for success. Thus, my suggestion is not to follow advice from your elders slavishly, but rather to always remember the importance of dancing to the beat of your own drummer. There is no single alchemist's formula for creating a successful scientist. Ultimately, you will feel better about your life in science if you live it on your own terms, follow your own scientific muses, have the courage to pursue the questions you really care about, and conduct science in the ways you think are most excellent. If you are fortunate, you will have superb mentors and colleagues along the way who will impart useful skills and valuable knowledge. They will also provide career advice and share with you their own trials and tribulations and paths to success. I suggest taking the latter with a grain of salt. The times are constantly changing, and yesterday's solutions may not work with tomorrow's problems. Most importantly, you cannot allow the advice of others to turn you into something that you aren't.

Brief Biography

I received my Ph.D. from Vanderbilt University in clinical psychology. My first faculty position was at Indiana University, and I joined the Department of Psychology at the University of California, Berkeley in 1986. I am currently the Director of the Institute of Personality and Social Research, am the past director of the Bay Area Predoctoral Training Program in Affective Science, and have served two terms as

director of Berkeley's Clinical Science Program. My research program is in the area of human emotion, studying the organization of physiological, behavioral, and subjective systems; the ways that these systems are impacted by neuropathology, normal aging, and culture; and the role that emotions play in the maintenance and disruption of committed relationships. I am the past President of the Society for Psychophysiological Research, past President of the Association for Psychological Science, and current President of the Board of Directors of the Psychological Clinical Science Accreditation System. I received the inaugural Mentor Award for Lifetime Achievement from APS in 2013, the William James award for Lifetime Achievement in Basic Research from APS in 2014, and the Distinguished Scientific Contributions to Psychophysiology from the Society for Psychophysiological Research in 2015.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cuthbert, B. N., & Insel, T. R. (2010). Toward new approaches to psychotic disorders: The NIMH Research Domain Criteria project. *Schizophrenia Bulletin*, *36*, 1061–1062.

Daniels, R. J. (2015). A generation at risk: Young investigators and the future of the biomedical workforce. *Proceedings of the National Academy of Sciences, USA*, *112*, 313–318.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Erlbaum.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., . . . Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010. *PLoS Medicine*, *10*, e1001547. doi:10.1371/journal.pmed.1001547

Ferrer, E., & Song, H. (2012). Longitudinal structural models for assessing dynamics in dyadic interactions. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 601–616). New York, NY: Guilford Press.

Graham, S. (1992). Most of the subjects were white and middle class: Trends in published research on African Americans in selected APA journals, 1970-1989. *American Psychologist*, *47*, 629–639.

Insel, T. R. (2015). The NIMH experimental medicine initiative. *World Psychiatry*, *14*, 151–153.

Insel, T. R., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, *167*, 748–751.

Insel, T. R., & Gogtay, N. (2014). National Institute of Mental Health clinical trials: New opportunities, new expectations. *JAMA Psychiatry*, *71*, 745–746.

Jost, J. T., & McGuire, W. J. (2013). An additional future for psychological science. *Perspectives on Psychological Science*, *8*, 414–423.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 617–627.

Levenson, R. W. (2014). The future of clinical science training: New challenges and opportunities. *Psychological Clinical Science*, *2*, 35–45.

Levenson, R. W. (in press). Clinical psychology training: Accreditation and beyond. *Annual Review of Clinical Psychology*.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561–614). New York, NY: Plenum Press.

National Institute of Mental Health. (1999). *Report of the National Advisory Mental Health Council's behavioral science workgroup: Translating behavioral science into action*. Washington, DC: Author.

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.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 1–8.

Paul, G. L., & Shannon, D. T. (1966). Treatment of anxiety through systematic desensitization in therapy groups. *Journal of Abnormal Psychology*, *71*, 124–135.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.

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

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.