

Psychology's Replication Crisis as Scientific Opportunity: A Précis for Policymakers

Policy Insights from the
Behavioral and Brain Sciences
2018, Vol. 5(1) 134–141
© The Author(s) 2017
DOI: 10.1177/2372732217749254
journals.sagepub.com/home/bbs



Joseph Lee Rodgers¹ and Patrick E. Shrout²

Abstract

Psychological science is in the midst of what has been referred to as a “replication crisis.” The realization that many individual findings do not replicate in new studies has led to questioning the scientific method and the integrity of psychological science. We review the history of the replication crisis, and its positive and negative effects. Most of the elements of the replication crisis are re-emergent issues that methodologists have studied in the past, but to which researchers have become increasingly sensitized. Ultimately, we argue the value of the replication crisis, in that it has led to positive self-examination within our science and to the development of new and innovative methodology. The field is emerging from the replication crisis with a realization of the importance of multiple replication efforts, and an improved ethic of openness and transparency in the conduct of research.

Keywords

replication crisis, methodology, policymakers, NHST, meta-analysis, Bayesian methods

Tweet

Psychological science has turned the replication crisis into an opportunity to strengthen psychological scientific claims.

Key Points

- The realization that many individual findings do not necessarily replicate in new studies has led some observers to perceive a crisis that questions the scientific method and the integrity of psychological science.
- Behavioral research is naturally probabilistic, not deterministic; probabilistic results will always lead to failures to replicate, with predictable patterns, even when an effect is present.
- Psychological science should focus on overall patterns of replication results from multiple studies rather than single efforts to replicate a finding.
- The recent replication crisis has led to improvements in scientific procedures and conventions that will make the production of new knowledge more efficient and more effective.
- Psychological researchers must embrace an ethic of openness and transparency in the conduct of research.

Introduction

Oxford English Dictionary: A crisis is “a state of affairs in which a decisive change for better or worse is imminent.”

Vocabulary.com: “Crisis . . . a difficult . . . time in which a solution is needed.”

Psychological science is in the midst of what many refer to as a “replication crisis.” The purpose of this article is to define, explore, and interpret the replication crisis for those who are interested in its general implications. By its very definition, a crisis requires some kind of change or solution. What changes are needed? Are the needed changes specific to psychology or to science in general?

We will argue that psychological research does indeed have some problems; like any science, it is imperfect. We also argue these problems are not new. Indeed, for many statisticians and methodologists, the rediscovery and deeper examination of known problems does not itself warrant declaring a crisis. However, having these problems reported in the general press and being discussed in scores of articles in psychological journals and sessions of scientific meetings makes it clear that many people think that change in scientific conventions and evidence is needed. Treating those challenges immediately and carefully at this time is both

¹Vanderbilt University, Nashville, TN, USA

²New York University, New York City, USA

Corresponding Author:

Joseph Lee Rodgers, Department of Psychology and Human Development, Peabody College of Education and Human Development, 230 Appleton Place #552, Vanderbilt University, Nashville, TN 37203-5721, USA.
Email: joseph.l.rodgers@vanderbilt.edu

important and necessary. It can improve research psychology in many ways. In this introduction, we comment on the nature of scholarly crises, both generally and in relation to the current replication crisis in psychological research.

Crises are intrinsically social events. In this article, we recap how this crisis was defined and how it is evolving. As a social event, it affects some groups more than others. For example, the replication crisis is being experienced more by social psychologists than by developmental, cognitive, or quantitative psychologists because more findings from social psychology are being questioned than from other areas. However, declaring a crisis can create opportunities for meaningful change. Crises can also create situations of disillusionment and distrust. Understanding can clarify what led to the proposal that a crisis exists.

Why is it important to dwell on the replication crisis, especially if it is primarily limited to a few subfields of psychology? We provide three answers. First, the replication crisis is an emergent movement within the field of psychological research (relevant to many other fields, including other social sciences and medial/biological sciences) that has and will continue to influence *how such research is conducted*; many sophisticated thinkers outside the professional research community are interested in psychological (and broader) research. Second, the replication crisis signals new and increased attention to the “sociology of psychological science,” the part of research that involves collaboration; focuses on ethics, scientific norms, and social structures within the profession; and quickly becomes visible and interesting to those outside the research profession. Third, some (not all) psychological research has obvious and substantial implications for developing, delivering, and implementing *policy* in government, education, and industry. Psychological research can help reveal best practices for public school education. Psychological research can help politicians understand how their leadership style will be received by their constituents. Psychological research can help business leaders become more effective communicators with their employees. However, if the replication crisis casts a pall on results from psychological research, its utility will be diminished.

We explain and interpret the replication crisis to provide perspective to understand the background and likely outcomes of the crisis. Our intended audience includes those both inside and outside of psychology. Several other articles with a similar goal can be consulted for their own interpretations of the crisis, each directed to a slightly different audience of external observers: see Ferguson (2015); Jasný, Wigginton, and Watts (2017); Lilienfeld (2017); and Pashler and de Ruiter (2017).

History and Description of the Replication Crisis

In a recent article in the *Annual Review of Psychology* (Shrout & Rodgers, in press), we argued that the perception

of a replication crisis was fueled by three specific events. First, in a short period of time, there occurred several fraud/misconduct cases, with particular relevance to social psychology. Second, a new and memorable name was given to scientific impropriety, Questionable Research Practices (QRPs; see Simmons, Nelson, & Simonsohn, 2011). QRPs reference both old issues (e.g., hypothesizing after the results are known, or HARKing, see Kerr, 1998; and problems with publication bias, see Rosenthal, 1979) and new issues (e.g., voodoo correlations, see Vul, Harris, Winkielman, & Pashler, 2009; and suggestions to change the traditional α -level from .05 to .005, see Benjamin et al., 2017). Third, a research effort by the Open Science Collaboration (OSC; 2015) found that only 36 of 100 replication results matched the original study.

In fact, many other events and publications contributed to the sense that a crisis was occurring, including Ioannidis's (2005) assertion in the medical literature that results hardly ever replicate, and several full issues of major research journals devoted to discussions, examination, and dissection of the newly identified crisis. However, many of the methods and conventions that were questioned in the crisis-fueled discussion had been questioned for decades. In this sense, the exact timing of the replication crisis is vague; we can argue, in fact, that it has been in development since the early 20th century, when the theory of Null Hypothesis Significance Testing (NHST) was developed. NHST is the statistical framework that has been employed within virtually all of social, behavioral, medical, and biological science for almost a century. Its legitimacy relies on the assumption that scientific results are subject to consistent and constant replication.

If the seeds of the replication crisis were planted around a century ago, the germination has continued ever since, and crises with similarities to the recent one have erupted during certain historical periods. Around 50 years ago, Milgram's (1963) obedience studies and Zimbardo's (1971) prison experiment were especially controversial, and caused psychology to change many research practices. Around 25 years ago, another controversy began over the legitimacy of NHST. Cohen (1994) was particularly critical of NHST, and Schmidt and Hunter (1997) issued a call to “outlaw” NHST. The strength of opinions and fervor then at least matched current interest in the replication crisis. Most who worked (and lived) through that earlier “methodology crisis” believe that the field was substantially improved by the self-examination and methodological shifts that occurred (see, in particular the task force report by Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999, and articles in the edited book by Harlow, Muliak, & Steiger, 1997).

However, the revisiting of these issues in the context of the most recent crisis has led to new insights and formulations of solutions. For example, there exist two slightly different problems, one called “reproducibility” and another

concerning “replication.” Whether a result is reproducible raises the issue of whether an analysis of exactly the same data would arrive at exactly the same results and conclusion. An example of reproducibility is the challenge for computer scientists who report algorithms that produce different results when implemented on different platforms (e.g., Sandve, Nekrutenko, Taylor, & Hovig, 2013). Whether a result is replicable raises the issue of whether a different research team attempting to run the same study and collecting *new and different data* would find the same results and conclusions.

The concept of replication is at the basis of the NHST paradigm that was developed in the 1920s by Ronald Fisher, Jerzy Neyman, and Egon Pearson. That paradigm has been the foundation for statistical analysis used in many different disciplines, and in psychology in particular. The NHST paradigm has been criticized by scholars for more than half a century (see Rodgers, 2010, for a review). The particular set of criticisms that has brought the replication crisis to a head is the concern that many results in the psychological literature cannot be replicated by careful research teams following up an interesting published finding; in fact, in some cases, the results are not even reproducible when the original data are made available.

The OSC (2015) raised this concern most vividly. Published findings are expected to replicate most of the time when other teams attempt to reproduce the original research. But in the OSC study, only 36% matched the original finding. What this means depends partly on balancing the risk of two kinds of errors, called Type I and Type II errors. A Type I error is a *false positive*, which occurs when a researcher claims to find an effect that is not really there. A Type II error, however, is a *false negative*, which occurs when a researcher fails to find an effect that exists in nature. Researchers attempt to keep Type I errors (false positives) at a fixed level, which is called the α level, usually set at or below 5%. The OSC finding was interpreted to mean that there are many more Type I errors in the published literature than there should be.

Various opinions have been expressed about what proportion of reported findings are Type I errors, but this quantity cannot be known from OSC results, as has been discussed by a number of statistical experts. In particular, Maxwell, Lau, and Howard (2015) suggested that virtually all of the 100 replication efforts had objectionably low statistical power. Statistical power is the probability that a true effect of a certain size would be detected by a specific study design. Maxwell and his colleagues argued that the sample sizes of the OSC replication studies were too small to achieve reasonable statistical power. If power is low, it is no surprise that only 36% replicated. The OSC group thought they had designed studies with adequate power, but Maxwell and his colleagues showed that their calculations resulted in substantial underestimates.

Consequences of the Replication Crisis

The recent replication crisis has had both positive and negative consequences (Shrout & Rodgers, in press). On the positive side, the replication crisis has generated new methodological developments (e.g., replication teams, Bayesian applications, new meta-analytic approaches, new mechanisms for registering predictions of proposed studies), and also deeper understanding of some basic methodological principles (e.g., more nuanced understanding of statistical power, the role of the p value, and the implications of editing of nonsignificant findings out of reports). The crisis has forced psychology to address some long known problems, and to provide at least partial fixes for those problems (e.g., thoughtless over-reliance on $\alpha = .05$; emphasis on statistical rules-of-thumb in settings that do not justify those; chronically underpowered studies). In Shrout and Rodgers (in press), we provided a list of recommended steps to improve the quality and replicability of findings that are claimed from research studies (whether exploratory or confirmatory). Efforts to increase likelihood of replication should cause psychological science to mature more quickly, and to identify real findings more effectively, avoiding Type I errors.

The crisis has also had a number of negative effects. First, the reputation of psychological research has suffered, both inside and outside the field itself. Granting agencies that fund psychological science and R&D groups within industry may well have concerns over QRP's and similar issues, and this could mean that obtaining funding for important research is more difficult after the crisis (though this is not documented, and is only speculation; see Jasny et al., 2017, and Lilienfeld, 2017, for broader discussion of the relationship between the grant culture and the replication crisis).

Second, some of the proposed fixes of the QRP problems impede rather than promote quality research. One example is proposed rules-of-thumb that are supposed to handle all methodological challenges of a certain type, without recognition of the fact that there are different values and requirements across disciplines (or even across research teams within a discipline). To illustrate, most scholars would accept that the cost-benefit trade-off associated with false positives and false negatives should be tuned differently for researchers developing a cancer drug compared with social psychologists studying the effects of priming.

A third negative effect of the replication crisis is that requirements of increased sample sizes and additional logistic requirements for the publication process may have differentially negative consequences for young scholars as they work to publish articles, get tenure, and obtain funding (e.g., Finkel, Eastwick, & Reis, 2015). Although economic forces in the academic market are likely to self-correct over time, if young scholars perceive that the potential for publication and promotion are indeed less supportive than they expected, negative short-term effects indeed can be predicted.

A final negative effect of the replication crisis has been to reinforce in the minds of some professional researchers and some in the policy (and broader) communities that psychological research findings are either clearly correct or incorrect. The probabilistic nature of research—which emerged from Fisher and Neyman–Pearson theories of randomness and null effects—mitigates against the view that a finding should either be replicated or not, in some ultimate sense. Given an original, statistically significant finding, and a dozen contradictory replication efforts, there is still a non-zero probability that the original finding is correct and that subsequent failures to replicate findings are Type II errors (false negatives). The frequently implied expectation—not only by journalists but also by sophisticated researchers—that research findings should replicate in every new study reflects a misunderstanding of the nature of the research enterprise. As discussed in Shrout and Rodgers (in press), Fisher would have likely viewed the recent replication failures about which so much attention (and angst) has developed as “business as usual.”

Responses to the Replication Crisis

We are aware of two types of responses to the replication crisis. First, many psychological researchers have viewed the crisis through an “old lens,” and have developed responses that reaffirmed existing approaches. Second, some psychological researchers have developed new approaches in response to the crisis. We review those in relation to several replication crisis “highlights,” ones that have already been mentioned but which provide a context for looking at various responses to the crisis.

Responses Using Existing Approaches

We elaborate now on the results of the OSC (2015), whose efforts to replicate 100 studies yielded only 36 results that matched the original studies as being statistically significant. Each of these replication efforts was a traditional effort, in which the replication team attempted to collect new data by matching the design, administration, and analytic approaches from the original study. It is a long-standing and traditional approach to hold a second study’s findings up against the study that it (approximately) replicates. One of the positive developments of the replication crisis is the recognition of subtle challenges associated with this process. For example, the OSC study researchers made a number of slight to substantial revisions to the original studies, revisions that were required for pragmatic reasons, but ones that may have attenuated the apparent success of the replication. This challenge is discussed in detail by Stroebe and Strack (2014), who noted that “the myopic focus on ‘exact’ replications neglects basic epistemological principles” (p. 60).

Second, many past replication efforts used findings from the original effort to estimate statistical power required for a replication effort, a process that has consistently and substantially under-estimated the required power. Maxwell et al. (2015) demonstrated that power calculations need to take into account sampling variation of estimates of effect sizes when planning replication studies. In addition to the statistical point made by Maxwell and colleagues, bias in effect size determination results from the traditional requirement that only statistically significant findings can be published. One indication of these biases is that the OSC effect size estimates were about half the size of the original effects that were being replicated.

Third, many researchers have thought that the veracity of a finding can be checked with a single replication study. Indeed, this was the model used by OSC: 100 specific findings were replicated by a single second study. A key take-home point of the replication crisis is that knowledge building depends on patterns of replication results from multiple studies rather than single efforts to replicate a finding. An exciting development from the replication crisis was the initiation of the Registered Replication Report (RRR) mechanism by the Association of Psychological Science (Simons, Holcomb, & Speelman, 2014). To use the RRR mechanism, a potentially important finding that has already been published is proposed to an editor for replication and if accepted, multiple collaborative teams each carry out a replication study. The journal commits to publish a well-prepared RRR report regardless of findings, which are reported using meta-analysis techniques. The development of whole series of replication efforts, followed by statistical approaches in which all the results are combined into an overall meta-analysis, is a new and statistically defensible approach to handling replication variation (see Braver, Thoemmes, & Rosenthal, 2014).

Fourth, Etz and Vandekerckhove (2016) used a Bayesian method to provide support for the seriousness of the findings from the OSC study. Methodological expansion includes applying old methods in new ways and to new problems. A number of Bayesian applications have been stimulated by the replication crisis (e.g., Kruschke & Liddell, 2017; Wagenmakers et al., 2012). However, Gilbert, King, Pettigrew, and Wilson (2016) reran the OSC statistical analyses using their original traditional approach, and interpreted the findings as being more often supportive than successful replication had occurred than did the original OSC authors. More broadly, the explanations in this section are excellent examples of how the replication crisis has generated traditional responses, which themselves have led to improvements in methodological and statistical procedures.

Responses Involving New Developments

Among the most innovative of the new developments associated with the replication crisis is a broad initiative by Nosek

and colleagues (2015) called the Center for Open Science and the Open Science Framework (OSF), a not-for-profit organization that promotes transparency in research activity. In particular, OSF supports preregistration of hypotheses and study designs, as well as posting computer code and analysis syntax for examination and use by other researchers. Efforts such as the OSF can successfully address QRPs that involve taking advantage of chance outcomes and that mix exploratory and confirmatory analyses in inappropriate ways.

Another major positive initiative emergent from the Replication Crisis is increased attention to the variability built into effect sizes (see Shrout & Rodgers, in press). Furthermore, focusing on finding “sweet spots” in intervention studies that will render relatively large effect sizes will naturally increase statistical power and the ability to replicate. Once an effect is established in an ideal setting, its generalizability can be systematically examined.

Both the technical and the popular press have responded with interest (even fascination) to the replication crisis. When a community of scholars begins to self-criticize, and to challenge the legitimacy of the whole field, reporters have ample material from which to craft interesting human-interest stories. When famous scholars “take sides” and argue in print about issues of both methodology and professionalism, reporters can create interesting perspectives for readers.

Yet, science, and especially social-behavioral science, is by its very nature uncertain. Journalistic treatment has a difficult time dealing with uncertainty, however, consistently expecting definitive quotes from scientists and treating science as more deterministic than it is. Many of the nuances and requisite qualifying of findings recognized as essential by professional researchers—and which annoy journalists looking for a good quote—have been part of the treatment of the replication crisis. Furthermore, modern social media has multiplied the ability of both journalists and also the interested public to study and comment on scientific processes—including nonreviewed writing going quickly into print for thousands to read, followed by dozens or even hundreds of public comments arguing back-and-forth over important issues. The methodology crisis of the 1990s mentioned above was “vetted” on list-serves; the replication crisis has been both a product of, and also treated with great fascination, by those posting on Facebook, Twitter, blogs, and other social media. The effect of modern social media is to increase the speed with which important—and also contentious—issues can move into the public arena. As a result, the public is more aware of the replication crisis than it would have been in earlier times (and possibly more confused as well).

Take-Home Messages From the Replication Crisis

This section begins by asserting that psychological researchers do value, substantially, input from those outside the

arena of psychological research. However, most of those who are external to research are interested in *findings* from the psychological literature that are useful in application, and that are surprising. Even the more sophisticated general readers are less interested in the details of psychological methods (including design, measurement, and statistical analysis) than they are in the empirical findings and substantive interpretations. This means that our formulation of take-home messages needs to be general rather than tied to the technical advances most exciting to quantitative methodologists. We present three take-home messages of potential broad interest.

The Replication Crisis Has Led to Positive Growth in Psychological Science

As the initial section noted, designating the recent period of self-examination by psychological scientists as a crisis can be interpreted in different ways. Some might say the scientific sky is falling and that all of our psychology textbooks are filled with Type I errors. Others might say that all of the recent fuss is simply a recapitulation of old known problems, and point to the many advances in behavioral and brain science that have improved methods of learning, treatment of cognitive symptoms of depression, understanding of automatic and effortful information processing, and design of automobile control systems.

We take a position in between. The recent replication crisis did wake up many psychological scientists and alerted them to conventions in the research and publication practices that led to unnecessary false positive claims and a false sense of closure about published work. We have presented work on the replication crisis at major psychology conferences in standing-room-only settings; the excitement and positive sense of addressing important issues with the goal of moving forward have been palpable in those settings. A strong commitment to solve our general problems is much more important than the semantic issues involving “crisishood.” For many psychologists, this statement has characterized careers, and is an old commitment, which brings us to the second take-home message.

Most Elements of the Replication Crisis Are Not New

Most elements of the replication crisis are not new, and some are very old. The conceptual value of replication lies at the heart of Fisher’s statistical developments in the early 20th century (e.g., Fisher, 1925). Fisher almost assuredly would *not* have viewed the recent developments as anything other than “business as usual.” Findings have variability. We cannot predict any particular finding, but we can predict the variability across findings, and that is the challenge that the scientist and statistician engage when they conduct a

research study. Of course some of the QRPs (including HARKing and publication bias) are on the ethical edge (or across, in many cases), and require adjustment in our practices. Careful attention to the distinction developed by Tukey (1977) between exploratory and confirmatory statistical procedures can solve many (though not all) of the QRPs that have been identified.

Virtually, all the recently defined QRPs have existed for many years, with awareness, extensive discussion, and efforts to address them on a regular basis in past work. Preregistration was discussed in the authors' graduate classes in the 1970s, and likely before. Publication bias has been a concern since even before Rosenthal (1979) referred to the "file drawer problem." HARKing was named in the 1990s, but was far from a new (or newly recognized) phenomenon. Awareness of the challenges of underpowered studies has existed since the first half of the 20th century, when NHST was being developed.

We believe that outright fraudulence or deep corruption in research settings does occur, though at a fairly low level, and must be addressed and managed when it emerges. We also believe that other approaches closer to litigation rather than journal practices are necessary to handle such cases, and each must be resolved at the individual level. We also believe that the kind of attention to both ethical and pragmatic methodological issues that are emerging from the replication crisis can help to deter those inclined to cheat. We emphasize to policymakers that psychological research is not unusual in the rate of unethical practices, nor are we unique in being concerned to minimize such practices and to have systems in place to discourage their practice. We leave detailed discussion of such methods for other accounts.

The Replication Crisis Is a Positive Step Forward for Scientific Methodology

Lilienfeld (2017) said, "I see the replication crisis as among psychological science's finest hours" (p. 660). He held this opinion because it has led to researchers being properly self-critical in evaluating evidence and to developing methodological remedies to various problems. Similarly, in our *Annual Review of Psychology* article, we asserted that "recent attention to replication in particular and knowledge generation more generally has led to remarkable and positive effects" and that "the future of psychological science is bright" (Shrout & Rodgers, in press). Even if the majority of the problems that have been discussed during the replication crisis have been discussed previously, the social construction of the replication crisis and the intense debate has led to awareness of methodological issues that has eluded previous methodology advocates. The past warnings of Bakan (1966), Greenwald (1976), Cohen (1994), and others were noted by methodologists but largely ignored by the general research community.

To be clear, psychological research has produced scientific and methodological advances throughout its history, including such widely used techniques as factor analysis and structural equation modeling, measurement level, meta-analysis, multilevel modeling, mediation and moderation, item response theory, and many research design innovations. Those techniques are not likely to be impressive to those outside the research enterprise, but to researchers, they represent some of the most important methodological tools in psychology and related sciences. Obviously, we do not necessarily need a crisis to stimulate us to be creative and innovative methodologically.

However, the field does respond, actively, to criticisms from both within and without. Advances in using meta-analysis to improve the methodology of replications (e.g., Braver et al., 2014) have been a positive response to the replication crisis. Preregistration has been used for decades, yet the replication crisis has brought it further forward in importance. Standard statistical teaching and research approaches, such as the tuning of alpha as explicit control over false positives (and, therefore, implicit balancing of those against false negatives and statistical power) has improved, as researchers have been sensitized to their importance by the replication crisis. Even Bayesian statistics, the mathematically deep and challenging branch of statistics within which most of the field began (e.g., Zabell, 1989), has expanded in relation to replication science. Each of these, and other methodological procedures, have developed in relation to widespread concern over QRPs, HARKing, and other threats identified in the context of replication.

Another sense in which the replication crisis can ultimately be a positive event for psychological research and methodology would occur if the expansion and opening of research methods can symbolize before the general public that psychologists perform their research with a general orientation toward transparency and honesty. Our field is being criticized for weaknesses in this domain at the moment. A proper response—whether we feel that the criticism is justified or not—is to emerge as exemplary scientists who are both self-critical and responsive to external criticism, who are highly trained in both the ethics and practice of proper research methods and who implement that training, and who are committed to supportive and honest collaboration (i.e., active collaboration with our immediate colleagues, and broad support for the legitimate talent and good-faith efforts of those across the fields of social and behavioral science).

Policy Insights From the Replication Crisis

What does the replication crisis suggest for funding research in general, selecting topics that need further development, and evaluating progress of individuals and programs? For the most part, the replication crisis does not require major

policy adjustments regarding science generally, psychological science in particular, and applications of psychological science to education, management, industry, or health. We review here how we think those outside of psychological research can best support our discipline's immediate and future health.

First, those inside and outside the research arena should be patient with the scientific process of proposing advances, checking these proposals, and correcting mistakes iteratively. None of us should overreact to the portrayal of recent events as a crisis. As noted repeatedly, little within the replication storms is new, although the storms are themselves real and reflect honest problems and challenges. But it takes time, hard work, a spirit of openness, and honest self-evaluation to turn internal tension and negative self-evaluation in a positive direction. As veterans who have observed such crises before, we beg for patience from those making science and social engineering policy.

Second, we also suggest that those setting financial and institutional policies consider supporting methodology groups and centers who are working to define and institute reforms to reduce QRPs and other problematic practices identified by the replication crisis. None of us should underreact. The Center for Open Science is an example of a new institution that provides tangible support for registering hypotheses and analyses, and facilitating openness by posting data and analysis scripts. Other institutions need to facilitate crowd-sourcing of replication studies and to promote training materials that will lead to more sophisticated and more open analyses. Science journalists need support to write about successful refinements of scientific methods as well as striking cases of fraud or scandal.

Third, policies should develop that encourage what Kuhn (1962) called "normal science," in addition to science that promises paradigmatic shifts. Some colleagues criticize graduate students and junior faculty who participate in replication trials because this service takes away from more creative research. Policies need to support a balance between the refinement of others' scientific findings and creating new insights; both processes are critical for psychological science to be successful. These policies will have implications for setting funding priorities; for hiring, promotion, and retention of faculty and researchers; and for allocating journal space for well executed studies with negative findings.

Conclusion

Both research psychologists and the interested public must be realistic about the perceived failings of scientific method and practice in psychology and related sciences. External observers who have paid careful attention to articles in psychology, to articles in the popular press, and to social media might conclude that *most* research psychologists have

engaged in volitional ethical violations and have forgotten or chosen to ignore the careful methodological training they obtained in graduate school. These conclusions would be incorrect. Psychological science has developed and applied remarkable methodological innovations during the past century, and the recognition of replication problems has led to development of new innovations that will help advance psychology as well as related social sciences and medical research.

The recent replication crisis has been a social event that has successfully promoted change in conventions and scientific practices. The new developments related to more openness in science and reduction of QRPs fit within a strong tradition of methodological investigation of scientific methodology. The field and those making policy that affect psychological research should recognize that knowledge attainment requires multiple replication studies with both exact and generalized designs, rather than single studies that confirm or disconfirm a claimed result. Finally, we celebrate the trend toward openness in science, the preregistration of both confirmatory experiments and even plans for exploratory research, and the collaboration of teams of scientists who work together to estimate the magnitude of interesting effects being reported in psychological science.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 423-437. doi:10.1037/h0020412
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., . . . Johnson, V. E. (2017). Redefine statistical significance. *Nature: Human Behavior*. Advance online publication. doi:10.1038/s41562-017-0189-z
- Braver, S. L., Thoenmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, *9*, 333-342. doi:10.1177/1745691614529796
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003. doi:10.1037/0003-066X.49.12.997
- Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the reproducibility project: Psychology. *PLoS ONE*, *11*(2), e0149794.
- Ferguson, C. J. (2015). "Everybody knows Psychology is not a real science": Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *American Psychologist*, *70*, 527-542.

- Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology, 108*, 275-297. doi:10.1037/pspi0000007
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver & Boyd.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science." *Science, 351*(6277), Article 1037. doi:10.1126/science.aad7243
- Greenwald, A. G. (1976). An editorial. *Journal of Personality and Social Psychology, 33*, 1-7.
- Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. doi:10.1371/journal.pmed.0020124
- Jasny, R. R., Wigginton, N., & Watts, D. (2017). Fostering reproducibility in industry-academia research. *Science, 357*, 759-761.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*, 196-217.
- Kruschke, J. K., & Liddell, T. M. (2017). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-016-1221-4
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Lilienfeld, S. (2017). Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science, 12*, 660-664.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*, 487-498.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371-378.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*, 1422-1425. doi:10.1126/science.aab2374
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), Article aac4716. doi:10.1126/science.aac4716
- Pashler, H., & de Ruiter, J. P. (2017, September). Taking responsibility for our field's reputation. *APS Observer*. Retrieved from <https://www.psychologicalscience.org/observer/taking-responsibility-for-our-fields-reputation>
- Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist, 65*, 1-12.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638-641.
- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology, 9*(10), e1003285.
- Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37-64). Mahwah, NJ: Lawrence Erlbaum.
- Shrout, P. E., & Rodgers, J. L. (in press). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*.
- Simons, D. J., Holcomb, A. O., & Speelman, B. A. (2014). An introduction to registered replication reports at Perspectives on Psychological Science. *Perspectives on Psychological Science, 9*, 552-555.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Perspectives on Psychological Science, 22*, 1359-1366.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*, 59-71.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Vul, E., Harris, C., Winkelman, P., & Pashler, H. (2009). Puzzling high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*, 274-288.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*, 632-638.
- Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604. doi:10.1037/0003-066X.54.8.594
- Zabell, S. (1989). R. A. Fisher on the history of inverse probability. *Statistical Science, 4*, 247-256. doi:10.1214/ss/1177012488
- Zimbardo, P. (1971). *The Stanford prison experiment: A simulation study of the psychology of imprisonment*. Stanford, CA: Stanford Digital Repository, Stanford University.