aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2018, Vol. 13(4) 411–417

DOI: 10.1177/1745691617751884
www.psychologicalscience.org/PPS
SAGE

Implications of the Credibility Revolution for Productivity, Creativity, and Progress

Simine Vazire
Department of Psychology, University of California, Davis

Abstract

The credibility revolution (sometimes referred to as the "replicability crisis") in psychology has brought about many changes in the standards by which psychological science is evaluated. These changes include (a) greater emphasis on transparency and openness, (b) a move toward preregistration of research, (c) more direct-replication studies, and (d) higher standards for the quality and quantity of evidence needed to make strong scientific claims. What are the implications of these changes for productivity, creativity, and progress in psychological science? These questions can and should be studied empirically, and I present my predictions here. The productivity of individual researchers is likely to decline, although some changes (e.g., greater collaboration, data sharing) may mitigate this effect. The effects of these changes on creativity are likely to be mixed: Researchers will be less likely to pursue risky questions; more likely to use a broad range of methods, designs, and populations; and less free to define their own best practices and standards of evidence. Finally, the rate of scientific progress—the most important shared goal of scientists—is likely to increase as a result of these changes, although one's subjective experience of making progress will likely become rarer.

Keywords

credibility, transparency, scientific progress, creativity, productivity

There is no doubt that psychology—along with many other scientific disciplines—is in the midst of a period of intense self-examination and self-improvement. These activities are not unusual for science; indeed, self-correction is said to be one of science's distinguishing features (Kuhn, 1970). However, the current period of self-examination is more extreme than usual, and in psychology in particular, self-scrutiny has reached a fever pitch. This is evidenced by the many articles in both the scientific literature and the media about psychology's replicability problem and about psychology's innovative approaches to self-improvement (e.g., Button et al., 2013; Munafò et al., 2017; Spellman, 2015). Underlying these discussions are concerns about the credibility of scientific claims; replicability is but one indicator of credibility, and the current "crisis" goes far beyond concerns about replication failures. Moreover, *crisis* implies that we are at a loss for solutions, when in fact we have identified many ways to improve science's credibility (Nelson, Simmons, & Simonsohn, 2018). Thus, I will refer to these current events—both the problems and the solutions—as the "credibility revolution" (Angrist & Pischke, 2010).

In this essay, I explore what the credibility revolution means for productivity, creativity, and progress in psychology. I begin by reviewing the most salient changes that have been brought about by the credibility revolution. Then, I review some common concerns about the implications of these changes for psychological science and consider each of these concerns in turn. I conclude that the changes brought about by the credibility revolution are likely to hamper the rate of individual researchers' productivity, could have a negative or positive effect on creativity depending on how the changes are implemented and what is meant by creativity, and are likely to increase the rate of scientific progress.

Corresponding Author:
Simine Vazire, Department of Psychology, University of California, Davis, One Shields Ave., Davis, CA 95616
E-mail: svazire@ucdavis.edu

Changes Stemming From the Credibility Revolution

As I write this in late 2017, the credibility revolution is still in full force. It is difficult to tell from this vantage point which changes will take hold and what will come next. Nevertheless, there are several developments that seem to have gained traction (see editorials, e.g., Eich, 2014; Lindsay, 2015; Spellman, 2015; Vazire, 2016) and are likely to have a lasting effect on the field, in some form and to some degree.

Transparency

One major theme of the credibility revolution is the need for greater transparency in the research and publication process (Kidwell et al., 2016; Lupia & Elman, 2014; Nosek et al., 2015; Spellman, Gilbert, & Corker, 2017; Vazire, 2017). This includes better documentation and accessibility of procedures, materials, data, and analysis scripts. Transparency also includes thorough reporting of all studies run, data exclusions, and robustness to alternative analytic models—in other words, asking researchers to report all results rather than permitting undisclosed selective reporting.

Preregistration

Another proposed solution is to encourage preregistration of new studies (van 't Veer & Giner-Sorolla, 2016). Preregistration involves writing down one's method, analysis plan, and predicted results before collecting data and then making this plan publicly available. This allows researchers and readers to know which analysis decisions and results were predicted a priori and can be interpreted as confirmatory (or disconfirmatory). It also helps researchers avoid the pitfalls of having to choose which analysis to report after seeing the results of the analyses or trying to remember one's hypotheses after seeing what patterns emerge in the data.

A special case of preregistration is registered reports, which are manuscripts that present a full introduction, method section, planned analyses, and anticipated results (Chambers, 2013). These manuscripts are submitted for peer review before the data are collected, while changes can still be made to the method. If the manuscript passes the review process, the authors are given an "in principle acceptance," and the manuscript is published regardless of the outcome of the study.

Direct replication

The credibility revolution was fueled in part by large-scale replication attempts that failed to replicate various findings published in our field's top journals (Cheung et al., 2016; Donnellan, Lucas, & Cesario, 2015; Ebersole et al., 2016; Eerland et al., 2016; Harris, Coburn, Rohrer, & Pasher, 2013; Johnson, Cheung, & Donnellan, 2014; Klein et al., 2014; Klein et al., 2017; O'Donnell et al., 2018). Whatever else these findings mean, they highlight the difficulty in replicating even well-established effects. Although our field is still refining the standards for how replications should be conducted and reported, there is little doubt that we will be seeing more replications after the credibility revolution than we did before (Makel, Plucker, & Hegarty, 2012).

Higher standards of evidence

Perhaps the most controversial recommendation that has come out of the credibility revolution is the proposal that we raise our standards for the quantity and quality of evidence we require before drawing conclusions (e.g., Giner-Sorolla, 2012; Schimmack, 2012; Simmons, Nelson, & Simonsohn, 2011; Tullett, 2015). One example of this shift is the proposal by Benjamin and colleagues (2018) that we redefine *statistical significance* and adjust our alpha to .005. The argument for this change is that, in the absence of the changes recommended above (i.e., complete transparency, preregistration, and/or direct replication), our convention of using an alpha of .05 and interpreting statistically significant results as compelling evidence for our research hypothesis is too liberal.

Another example of the shift toward higher standards of evidence is the "constraints on generality" statement proposed by Simons, Shoda, and Lindsay (2017). According to this proposal, authors would be required to justify the generalizations that they claim can be made from their findings and explicitly state any anticipated limits to the generality of their findings.

More generally, the credibility revolution has led to a push for greater emphasis on methodological rigor and a demand for better evidence to support strong claims (Bakker, van Dijk, & Wicherts, 2012; Munafò et al., 2017; Smaldino & McElreath, 2016). This approach emphasizes not just increased statistical rigor but also increased emphasis on using rigorous methods. In short, conclusions should be better calibrated to the quantity and quality of evidence supporting the claim.

Implications of Changes

Productivity

Will the credibility revolution lead to more or less productivity? Some have expressed the concern that the changes described above will lead to "slower science";

researchers will conduct fewer studies and publish fewer manuscripts per year. I anticipate that this will indeed be a consequence of these changes. It is important to acknowledge this possibility and to think about the consequences that this would have for our field and the people in it.

The credibility revolution will lead to slower science primarily because of the fourth change described above—increased standards for rigor. If we raise the standards for publication, researchers will have to put more time, effort, and resources into each publication, which will mean publishing fewer manuscripts.

Other changes described above may actually counteract this effect. For example, transparency, including the sharing of data and materials, would increase the ease with which researchers can test their research questions, by reusing others' materials or data. Indeed, open data may mean that we as a field get much more information out of each data set than we have in the past, which could represent a significant increase in productivity. In addition, the registered reports model of publication guarantees that all sound studies that are conducted in this framework will be published without having to successively submit one's manuscript to multiple journals. Finally, direct replications could make researchers more efficient by identifying false-positive effects quickly and openly, preempting a situation that some argue has been common in the past, wherein many independent laboratories spend months or years trying to replicate or build on an effect before giving up quietly and moving on.

I suspect that the productivity-dampening effects of higher standards will outweigh the productivity-enhancing effects of the other changes brought about by the credibility revolution. However, this is an empirical question, and the answer will likely vary across subfields and research frameworks.

What is the ideal level of productivity? An important question to ask is whether productivity should be a goal in its own right. Of course, wanting to be as productive as or more productive than other scientists is, and will remain, an important motivation for researchers. Individual differences in productivity (i.e., relative productivity) will always be of interest to hiring, promotion, award, and grant committees, and so, one's productivity relative to one's peers will always be of concern to researchers. Moreover, a researcher's relative productivity (i.e., being more productive than one's peers) has been shown to be a strong predictor of impact and contribution in science and other disciplines (Simonton, 1997, 2003, 2009). However, that does not mean that lower absolute levels of productivity (i.e., shifting the distribution of researchers' productivity, without necessarily affecting the rank order) are necessarily problematic. What is the value of high absolute levels of productivity?

One possible answer is that greater productivity means greater scientific progress (i.e., faster accumulation of knowledge). However, the connection between productivity and progress is not straightforward. Progress depends on the amount and quality of evidence presented in each publication. If high productivity is coming at the expense of quality or quantity of evidence per publication, maximizing productivity will not maximize—and could even undermine—progress, as discussed in greater detail below.

To some extent, there is necessarily a trade-off between valuing productivity and maximizing the quality of each individual research product. If too much emphasis is placed on producing more research, that research will be scrutinized less carefully (because researchers will be too busy writing up new manuscripts to take a close look at each other's—or their own—reports). Moreover, if too much of a premium is placed on quantity of publications, researchers will have a strong incentive to carve up their data into the "least publishable unit," once again driving down the informational value of each publication.

Of course, the reverse is also true; we can put so much emphasis on maximizing and verifying the value of each publication that productivity grinds to a halt. Some fear that the credibility revolution will lead to such a state. To be sure, a balance should be struck between valuing productivity and assuring that enough time and effort are spent on the research that goes into each publication. If the credibility revolution does in fact lead to a decline in productivity, whether this is a positive or negative change depends on which side of this optimal balance point one believes we currently are.

If the changes spurred by the credibility revolution do reduce productivity, one potential negative side effect to consider is that this change will not be spread evenly across subfields, methods, and populations. Researchers studying questions that require intensive methods or hard-to-reach populations will face greater challenges when trying to meet new standards of rigor (Cialdini, 2009). Thus, comparing researchers' productivity across diverse methods, designs, or populations could disadvantage some groups of researchers. Of course, these disparities already exist, and for this reason, it is common to compare researchers with their peers, matched on these characteristics. However, new disparities may arise, and we should track these changes so that we can adapt our evaluation of researchers to these changes.

Creativity

Another common anticipated consequence of the changes spurred by the credibility revolution is that they will hamper researchers' creativity. There are several different possible meanings of creativity in this context, and

I have observed three common uses of the term *creativity* in the context of discussions about replicability and credibility problems in psychology. I will consider each one in turn: (a) the riskiness of research questions; (b) the breadth of methods, designs, and populations represented in research; and (c) the freedom to choose one's preferred research practices and standards.

Riskiness. One possibility is that researchers will study only "safe" research questions and will be less likely to go out on a limb and propose a far-fetched theory that could be potentially transformative. Certainly, increased transparency and preregistration will make it clearer when a counterintuitive or extraordinary finding was unexpected, making it harder to make strong claims based on serendipitous discoveries. This may make people more hesitant to pursue these risky leads. Similarly, higher standards of evidence will likely mean that bold claims that have yet to be corroborated will need to be flagged as speculation. Together, these changes may put a damper on researchers' enthusiasm for pursuing creative, high-risk research projects. It is likely that part of the motivation to pursue such projects has been the positive reactions to bold claims (e.g., grant funding, recognition from colleagues, media attention). Constraining researchers' claims to be better calibrated to the strength of the evidence will make such projects less rewarding.

In my view, this would be a positive development; there are currently many rewards for drawing bold conclusions on the basis of weak evidence (e.g., publication in prestigious journals, media attention) and few costs (e.g., little chance of direct replication attempts). I am in favor of shifting this balance in a way that would make these bold claims harder to publish. This would necessarily make the published literature less exciting (fewer bold, surprising claims), but I believe sacrificing this type of creativity for greater accuracy is worth it (even if it did almost kill the field of personality psychology, according to Baumeister, 2016).

Diversity of methods, designs, and populations. Another possibility is that the changes brought about by the credibility revolution will lead to a single-minded focus on quantity of evidence (e.g., large samples, low *p* values, large Bayes factors) at the expense of quality of evidence (e.g., rigorous methods and designs, diverse samples and settings that allow for generalization). In other words, to collect as much data as possible, researchers may reduce the breadth of methods they use and populations they sample from. The issue here is not directly one of creativity, but it concerns the willingness to use the tools most appropriate for the research question, which sometimes requires a willingness to think outside the box.

This is a real risk. During the first few years of the credibility revolution, much attention was focused on the problems with small samples (Button et al., 2013; Fraley & Vazire, 2014), and there was a strong push to increase statistical power or precision—that is, to increase the quantity of evidence necessary to make scientific claims. However, in recent years, the focus has been more balanced, including discussions of the importance of both quantity and quality of evidence (Munafò et al., 2017; Vazire & Lucas, 2017). Furthermore, the credibility revolution has spurred efforts to aid in implementing intensive methods and collecting data from diverse samples (e.g., Chartier, 2017; StudySwap, 2017).

Ultimately, trends in the diversity and rigor of research designs and methods can and should be tracked empirically. It is easy to speculate about whether the credibility revolution has led or will lead to overreliance on convenient but flawed methods, but we should be wary of any such speculation when the evidence is easy to collect. It is probably still too early to know (because it takes a few years for research to go from the design and data collection stages to publication), but early signs suggest that there has not been a decrease in the use of intensive methods or diverse populations (Vazire & Lucas, 2017). However, these same data also suggest that the use of creative, diverse, and intensive methods was already rare in psychology before the credibility revolution. This is consistent with previous calls for greater attention to the overreliance on convenience samples (Henrich, Heine, & Norenzayan, 2010) and methods (Chandler, Paolacci, Peer, Mueller, & Ratliff, 2015). Perhaps the credibility revolution presents an opportunity to bring more attention to these issues. Indeed, calls for greater methodological rigor—including the request that researchers use the best methods available rather than the most convenient—are in line with the shift toward greater skepticism and better calibration between evidence and conclusions.

Freedom to choose one's research practices and standards. One possible consequence of the credibility revolution is that there will be more uniform norms and standards about what constitutes acceptable—and unacceptable—research practices. Baumeister (2016) referred to the "intuitive flair" that it takes to design and execute a study that can detect an effect with 10 observations per condition and lamented the possible demise of this feature of research. In a different but related vein, Coan and Finkel, in an episode of Coan's (2017) podcast, discussed the potential negative effects of imposing strict rules about research practices, and Coan expressed the view that he prefers to "let a thousand flowers bloom" and let each researcher decide how best to approach the research question. It is hard to disagree with this position in the abstract; researchers should have some freedom to choose the approach they think is best, so long as they

can justify their decisions to the satisfaction of the scientific community.

It is clear that the credibility revolution has led to a push for clearer (and higher) standards, which would necessarily mean less freedom for researchers to choose their own standards and best practices. Why should we tolerate this loss of freedom? What are the costs and benefits?

On one hand, if science is inherently self-correcting, letting researchers have the freedom to choose any research practice could be a viable approach to science. If postpublication critiques are common and given a good deal of weight, the flawed approaches will be weeded out and only findings based on rigorous practices will gain traction. However, the credibility revolution was ignited precisely because of a perceived failure in this self-correction process. Thus, the danger of giving researchers complete freedom to choose their own standards and norms is that there is no effective mechanism for ensuring that only findings based on rigorous methods will eventually become established fact.

If we accept that the status quo in psychology prior to the credibility revolution was not sufficiently effective at discriminating (either before or after publication) between robust findings and flimsy ones, then we must make one of two changes. We must either impose more constraints on what are considered acceptable research practices before a manuscript is accepted for publication or design a more effective system of postpublication evaluation and self-correction. There are conflicting views within the community of scholars who are concerned about replicability regarding which approach is more viable (Nelson, Simmons, & Simonsohn, 2012; Nosek & Bar-Anan, 2012). Moreover, neither solution seems popular with those who are not convinced that science has a credibility problem. Imposing stricter standards before publication is seen as impinging on researchers' freedom, whereas calling out flimsy results after publication is distasteful to those who are concerned about incivility in scientific discourse. But we cannot escape both. There is no way to make science self-correcting without either imposing strict standards when deciding what to publish or engaging in critical discussions of published work. We must choose at least one of these if we want to identify and correct false scientific claims.

Progress

The common goal among all scientists is to accumulate knowledge. Much of the heat surrounding discussions about credibility can be attributed to differences in opinion about the best way to maximize this progress. Those in favor of the changes brought about by the credibility revolution believe that they will lead to faster accumulation of reliable knowledge. Those skeptical of these changes believe that they will slow down progress. As philosophers of science have demonstrated, there is not a simple answer to this debate.

It is heartening to remember that we share this goal. Starting from this common ground, the next step is to ask whether we can name more specific values or goals that are shared across the scientific community. In other words, what conditions can we agree are necessary for scientific progress? One potential answer is the commitment to self-correction. All scientists acknowledge that error is inherent to the scientific process and that in order to accumulate knowledge, it is critical that we have mechanisms in place to detect and correct flukes and errors. Thus, we can safely assume that enabling self-correction is a core value shared by all scientists.

What are the conditions necessary for self-correction? Merton's (1942/1973) norms provide a compelling set of guiding principles. Although Merton developed these norms as a description of what he believed scientists actually do, recent research suggests that most scientists do not believe that their fellow scientists adhere to these norms (Anderson, Ronning, DeVries, & Martinson, 2010).

Merton's first norm, *universalism*, states that scientific claims should be evaluated on their own merit and not on the basis of the status of the person making the claim. This is a necessary condition for self-correction. Without this value, it would be very difficult to challenge the claims of high-status scientists. Universalism encourages scientists to question authority, question received wisdom, and evaluate all scientific claims purely on scientific criteria. Many of the changes spurred by the credibility revolution help achieve the goals of universalism. For example, increased transparency helps provide all scientists access to the information they need to critically evaluate each other's claims without having to request permission or make any accusations. This makes it easier for lower status scientists to scrutinize the claims of high status colleagues.

Merton's second norm, *communalism* ("communism" in the original), states that science and the basis for scientific claims should be open and available. Once again, this is a necessary condition for effective self-correction because without transparency, it is harder to detect errors. Clearly, many of the changes spurred by the credibility revolution are directly aimed at achieving the goals of communalism.

Merton's third norm, *disinterestedness*, states that researchers should be expected to seek the truth rather than self-interest. Clearly, humans are susceptible to self-interested motivations, but endorsing this norm and aspiring to disinterestedness are important for scientific

self-correction because if we accept this norm, we accept that correcting the scientific record is more important than protecting the reputation of an individual scientist. Of course, gratuitous harm to a researcher's reputation should not be tolerated, but when the reputational harm is an unavoidable side effect of scientific self-correction, the norm of disinterestedness states that those committed to science must be willing to put scientific accuracy above the self-interest of individual researchers. Once again, this is in line with many of the changes brought about by the credibility revolution (e.g., more direct replications, which could lead to the uncomfortable realization that published and cherished findings are not robust; Zwaan, Etz, Lucas, & Donnellan, 2017).

Finally, Merton's fourth norm, *organized skepticism*, states that scientists should engage in critical evaluation of each other's claims and that nothing should be considered sacred. Scientific self-correction relies on this norm because theories or findings that are treated as sacred cannot be corrected. Thus, the push for higher standards of evidence, and for practices such as preregistration, transparency, and direct replication that make it harder to (intentionally or not) exaggerate the evidence for an effect, is in the spirit of the Mertonian norm of organized skepticism. Self-correction requires being willing to put theories and effects to severe tests and accepting that if they do not pass these high bars, we should be more skeptical of them.

In a sense, our desire for scientific progress presents a conundrum. We can achieve the feeling of progress by giving our theories and predictions the best possible chance of success. For example, without transparency, preregistration, direct replication, or strict standards of evidence, researchers can make stronger claims (e.g., claim that an unexpected finding was predicted, or avoid finding out whether it can be directly replicated). If others accept these strong claims uncritically, we can tell the world about our achievements and likely get respect and gratitude from scientific and nonscientific communities. However, this temptation should be resisted if we want to maximize progress in the long run. To accumulate knowledge and build on robust findings, we must make it harder to achieve the feeling of progress in the short run.

Action Editor

Robert J. Sternberg served as action editor and editor-in-chief for this article.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

References

Anderson, M. S., Ronning, E. A., DeVries, R., & Martinson, B. C. (2010). Extending the Mertonian norms: Scientists' subscription to norms of research. *Journal of Higher Education*, *81*, 366–393.

Angrist, J. D., & Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*, 3–30.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543–554.

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*, 153–158.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., . . . Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*, 6–10. doi:10.1038/s41562-017-0189-z

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews: Neuroscience*, *14*, 1–12.

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, *49*, 609–610.

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science*, *26*, 1131–1139. doi: 10.1177/0956797615585115

Chartier, C. (2017, December 3). *The Psychological Science Accelerator: A distributed laboratory network*. Retrieved from https://osf.io/93qpg/

Cheung, I., Campbell, L., LeBel, E. P., Ackerman, R. A., Aykutoğlu, B., Bahník, Š., . . . Yong, J. C. (2016). Registered replication report: Study 1 from Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspectives on Psychological Science*, *11*, 750–764. doi:10.1177/1745691616664694

Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, *4*, 5–6.

Coan, J. A. (Producer) (2017, September 18). *Circle of Willis (Episode 2)* [Audio podcast]. Retrieved from http://circleofwillispodcast.com/episode-2-eli-finkel

Donnellan, M. B., Lucas, R. E., & Cesario, J. (2015). On the association between loneliness and bathing habits: Nine replications of Bargh and Shalev (2012) Study 1. *Emotion*, *15*, 109–119.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., . . . Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82. doi:10.1016/j.jesp.2015.10.012

Eerland, A., Sherrill, A. M., Magliano, J. P., Zwaan, R. A., Arnal, J. D., Aucoin, P., . . . Prenoveau, J. M. (2016). Registered replication report: Hart & Albarracín (2011). *Perspectives on Psychological Science*, *11*, 158–171. doi:10.1177/1745691615605826

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*, 3–6.

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, *9*(10), Article e109119. doi:10.1371/journal.pone.0109019

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, *7*, 562–571.

Harris, C. R., Coburn, N., Rohrer, D., & Pasher, H. (2013). Two failures to replicate high performance goal priming effects. *PLOS ONE*, *8*(8), Article e72467. doi:10.1371/journal.pone.0072467

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral & Brain Sciences*, *33*, 62–135.

Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence morality judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology*, *45*, 209–215.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., . . . Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), Article e1002456. doi:10.1371/journal.pbio.1002456

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "Many Labs" replication project. *Social Psychology*, *45*, 142–152. doi:10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., . . . Friedman, M. (2017). *Many Labs 2: Investigating variation in replicability across sample and setting*. Unpublished manuscript. Retrieved from https://osf.io/8cd4r/

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, *26*, 1827–1832.

Lupia, A., & Elman, C. (2014). Openness in political science: Data access and research transparency. *PS: Political Science & Politics*, *47*, 19–42.

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.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago, IL: University of Chicago Press. (Original work published 1942)

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, Article 0021. doi:10.1038/s41562-016-0021

Nelson, L. D., Simmons, J. P., & Simonsohn, U. (2012). Let's publish fewer papers. *Psychological Inquiry*, *23*, 291–293.

Nelson, L. D., Simmons, J. P., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*, 17.1–17.24.

Nosek, B., Alter, G., Banks, G., Borsboom, D., Bowman, S., Breckler, S., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*, 1422–1425. doi:10.1126/science.aab2374.

Nosek, B. A., & Bar-Anan, T. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, *23*, 217–243.

O'Donnell, M., Nelson, L., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., . . . Zrubka, M. (2018). Registered replication report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science*, *13*, 268–294. doi:10.1177/1745691618755704

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*, 551–566. doi:10.1037/a0029487

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*, 1123–1128.

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, *104*, 66–89.

Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psychological Bulletin*, *129*, 475–494.

Simonton, D. K. (2009). Varieties of (scientific) creativity: A hierarchical model of domain-specific disposition, development, and achievement. *Perspectives on Psychological Science*, *4*, 441–452.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), Article 160384. doi:10.1098/rsos.160384

Spellman, B. A. (2015). A short (personal) future history of Revolution 2.0. *Perspectives on Psychological Science*, *10*, 886–899. doi:10.1177/1745691615609918

Spellman, B. A., Gilbert, E. A., & Corker, K. S. (2017). *Open science*. Unpublished manuscript. Retrieved from https://osf.io/preprints/psyarxiv/ak6jr/download?format=pdf

StudySwap. (2017, September 28). *StudySwap: A platform for interlab replication, collaboration, and research resource exchange*. Retrieved from https://osf.io/view/studyswap/

Tullett, A. M. (2015). In search of true things worth knowing: Considerations for a new article prototype. *Social and Personality Psychology Compass*, *9*, 188–201.

van 't Veer, A., & Giner-Sorolla, R. (2016). Pre-registration in social psychology: A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12.

Vazire, S. (2016). Editorial. *Social Psychological & Personality Science*, *7*, 3–7.

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology*, *3*(1), 1.

Vazire, S., & Lucas, R. E. (2017, May). *How meta-science can help guide journal policy*. Paper presented at the 29th meeting of the Association for Psychological Science, Boston, MA.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2017). Making replication mainstream. *Behavioral & Brain Sciences*. Advance online publication. doi:10.1017/S0140525X17001972