

# On the Right Side of Being Wrong: the Emerging Culture of Research Transparency

December 29, 2021



A few years ago, a group of researchers decided to model a rare level of honesty for their colleagues. Each of them agreed to disclose their doubts about one of their published findings. They called it the “Loss-of-Confidence Project.”

“Our main goal was to destigmatize declaring a loss of confidence in one’s own findings,” Julia Rohrer, a personality psychologist at the University of Leipzig in Germany and leader of the project, told the *Observer*. “Because of that, we also decided to focus on instances in which it is hardest to be open—when the central finding of a study is called into question because of a clear theoretical or methodological problem for which one has to assume responsibility.”

The initiative resulted in a collection of [13 self-corrections published in \*Perspectives on Psychological Science\*](#). Rohrer and her colleagues are hoping their article will motivate other researchers to reassess their work and, if necessary, disclose potential mistakes or shortcomings to the larger scientific community.

In the decade since researchers in psychological science began uncovering methodological flaws in many studies—igniting the field’s “replication crisis”—scientists like Rohrer have been laying a foundation for a new culture in the social and behavioral sciences. They’re creating a climate in which

researchers can share their data and materials so that others can review and spot potential methodological problems, acknowledge errors without killing their careers, and get published even when their studies yield negative results.

“Real behavioral change in improving research practices has begun in psychological science,” noted APS Fellow Brian A. Nosek, a leader in the field’s efforts to bolster scientific methods and reliability. “The field is leading the way in making change and evaluating whether those changes are having their intended effects on improving credibility.”

## Growing pains

Scientific psychology spent the 2010s in a period of intensive (and often rancorous) self-examination, as lab after lab reported smaller effect sizes or insignificant results when attempting to replicate some of the field’s most influential studies. Reformers pointed to the widespread use of results-skewing research practices like *p*-hacking (improper data selection or statistical analysis to make a nonsignificant result significant) and hypothesizing after finding a significant result. Authors of those studies found their reputations bruised if not battered.

Psychology is far from the only scientific discipline facing a replication problem. Fields ranging from economics to biomedicine are grappling with similar issues about research integrity and rigor. But psychological science has stood out for the breadth of its self-correction initiatives.

Nosek, a University of Virginia (UVA) professor, is at the forefront of those reform efforts. In 2013, he and quantitative psychologist Jeffrey Spies founded the nonprofit Center for Open Science to increase transparency and integrity in scientific research. In the Center for Open Science’s first project, they assembled 270 scientists to replicate the results of 100 peer-reviewed psychological studies. Overall, 97% of the original studies reported statistically significant *p* values ( $p < .05$ ), but only 36% of the replication studies found statistically significant results. Moreover, effect sizes on average were half of what was reported in the original studies (Open Science Collaboration, 2015). The results of that project fueled the field’s drive to advance reproducibility.

Additionally, a variety of grassroots efforts have emerged to foster a more robust science. Among them is the Psychological Science Accelerator, a worldwide network of labs aiming to reconduct experiments on a mass scale. In its [first major project](#), reported in 2021, the group found strong evidence—albeit with some nuances—supporting the findings of [a 2008 study showing that humans judge unfamiliar faces on trustworthiness and physical strength](#) (Oosterhof & Todorov, 2008).

### Related content from this issue:

- [The Grand Challenges of Psychological Science](#): APS members share what they perceive to be the biggest challenges to the field.
- [Exploring the Mysteries of Self and Consciousness](#): New explorations of the myriad ways consciousness influences human experience.
- [Fully Credited](#): A new model of “contributorship” addresses the marginalization of early-career researchers in scientific publications.

- [Up-and-Coming Voices](#): Previews of emerging research on methodology and research practices

## Acknowledging mistakes

Underlying these initiatives is a desire to erase the stigma surrounding scientific errors to help scientists understand that acknowledging a misstep doesn't stain their careers.

Rohrer said the inspiration for the [Loss-of-Confidence Project](#) came from social psychologist [Dana Carney \(University of California, Berkeley\), who in 2016 wrote](#) that she no longer believed in her earlier findings regarding the emotional, behavioral, and hormonal effects of power posing. She pointed to small effect sizes, a tiny sample size, and *p*-hacking (Carney, 2016).

Rohrer's team issued an open invitation for psychological researchers to submit statements about their loss of confidence in one of their previously published works. They also conducted an online survey and asked respondents whether they'd lost faith in a previously published finding and why.

"The idea behind the initiative was to help normalize and destigmatize individual self-correction while, hopefully, also rewarding authors for exposing themselves in this way with a publication," Rohrer and her colleagues wrote in the *Perspectives* article.

Participating researchers submitted statements describing a study they believed no longer held up. Five reported methodological errors, four acknowledged invalid inferences, and seven acknowledged *p*-hacking—in many cases due to a poor understanding of statistical methods.

Rohrer's team also gleaned 316 responses to their survey, capturing data from scientists at all career stages. Forty-four percent of the respondents reported losing trust in at least one of their findings. Of those, more than half said the finding was due to "a mistake or shortcoming in judgment," and roughly one in four took primary responsibility for the error. Respondents also admitted to questionable research practices.

Only 17% of those who lost confidence in their results said their concerns were a matter of public record, and most of those cases involved acknowledgement in articles, conference presentations, or social media posts that weren't directly linked to the original study report.

Asked why they didn't communicate their lost confidence in a finding, some respondents said they were unsure how to do so, while others felt that disclosing their concerns was unnecessary because the finding had attracted little attention. Some worried about hurting the feelings of coauthors, and others worried about how their disclosure would be perceived (Rohrer et al., 2021).

Rohrer and her colleagues want to promote a culture of self-correction so that researchers can disclose subsequently discovered problems with a study without fearing damage to their reputations and their work.

“Many researchers are concerned that such openness and humility may harm their career prospects,” she said. “I hope that for these people, the Loss-of-Confidence Project shows them that self-correction isn’t *that* big of a deal—everybody makes mistakes from time to time.”

### **Citations Lag Behind Replication Failures**

Although failed replications have thrown a string of landmark psychological findings into question over the last 10 years, citation patterns in the literature haven’t kept pace, new research suggests.

Scientists led by psychological researcher Tom E. Hardwicke of the University of Amsterdam culled four major multilab replications that contradicted original results. Hardwicke’s team then looked at patterns of citations after those failed replications and found only a small decline in favorable citations of the original research.

“Replication results that strongly contradict an original finding do not necessarily nullify its credibility,” [Hardwicke and his colleagues wrote in \*Advances in Methods and Practices in Psychological Science\*](#) (AMPPS). “However, one might at least expect the replication results to be acknowledged and explicitly debated in subsequent literature.”

In another study, social scientists at the University of California, San Diego, used Google Scholar to examine the citation patterns of studies in psychology, economics, and general sciences, both before and after other labs tried to replicate them (Serra-Garcia & Gneezy, 2021). The researchers found that studies that subsequently failed to replicate were by far the most likely to be cited—and only 12% of those citations even acknowledged the unsuccessful replication.

In their AMPPS article, Hardwicke and his colleagues acknowledged that some replication results have been challenged. Even that doesn’t justify ignoring the incongruity, they explained.

“Because this debate remains far from settled, ideally any favorable citation of the original studies should at a minimum be accompanied by a co-citation of the replication results and some discussion of the discrepant results.”

### **References**

Hardwicke, T., Szücs, D., Thibault, R. T., Crüwell, S., van den Akker, O. R., Nuijten, M. B., & Ioannidis, J. P. A. (2021). Citation patterns following a strongly contradictory replication result: Four case studies from psychology. *Advances in Methods and Practices in Psychological Science*, 4(3), 1–14. <https://doi.org/10.1177/25152459211040837>.

Serra-Garcia, M., Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, 7(21), eabd1705. <https://www.science.org/doi/10.1126/sciadv.abd1705>.

## Normalizing replications

Other scientists want to move replications beyond the success/failure paradigm and into an accepted part of scientific advancement.

“Replications can be strong indicators to other scientists, the public, and policymakers that things are working as they should,” [John E. Edlund \(Rochester Institute of Technology\) and his colleagues wrote in a new article for \*Perspectives on Psychological Science\*](#). “Because replications can ensure sound results and spark conversations about the research findings, they can propel science forward ... By valuing replications in our scientific communities, we normalize replications as part of the scientific process, allowing for beliefs to be modified as evidence emerges” (Edlund et al., 2021).

Fostering that outlook requires a change in the way scientific reputations are built, as Nosek, Rohrer, and a team of 14 other researchers suggested in a new [article in the \*Annual Review of Psychology\*](#). Many scholars have based their professional identities on their findings, which can lead them into a defensive stance when those results don’t replicate (Nosek et al., 2021). And past behavior on social media validates those concerns: Many researchers have seen their character and intentions attacked after others discover shortcomings in their work.

But a 2016 report out of UVA suggests that scientists who acknowledge a failed replication of their work may see their reputations bolstered rather than impugned. Psychological scientists Charles R. Ebersole (now at the American Institutes for Research) and Jordan R. Axt (now at McGill University) worked with Nosek to examine this phenomenon in an online survey of nearly 4,800 U.S. adults. Participants were shown descriptions of one hypothetical scientist who produces “boring but certain” results and another who produces “exciting but uncertain” findings. (For a subset of the sample, the words “reproducible” and “not very reproducible” were used instead of “certain” and “uncertain.”) The respondents were asked to rate each scientist’s intelligence, ethics, reputation, admirability, and employability—and those ratings showed that they preferred the scientist who produced more certain results over the scientist whose findings were more exciting.

Ebersole and Axt then presented participants with various scenarios involving scientists whose findings failed to replicate. The respondents viewed researchers’ ethics and ability more favorably when they acknowledged a failure to replicate compared to when they criticized it. Moreover, they gave the highest ratings on those dimensions to researchers who conducted replications of their own work, especially if they published an unsuccessful replication instead of dismissing it.

Ebersole, Axt, and Nosek conducted a similar survey with 313 psychological researchers and 428 college students and found comparable results. But unlike members of the general population, these participants said they thought a scientist who produced exciting but uncertain results would experience greater career success (Ebersole et al., 2016).

The survey findings underscore a call for changing the incentive structure in science. For years, researchers who want to improve scientific integrity have complained that journal editors, funders, and academic institutions favor researchers who produce novel findings, often at the expense of robustness. University administrators want faculty whose scientific records bring in grant funding. Grant applicants

exaggerate their expected findings to potential funders. And journal editors favor positive results, prompting researchers to bury null findings that likely won't get published.

“Sustainable change toward rigor and transparency requires that the reward systems for publication, funding, hiring, and promotion actively promote rigor and transparency,” Nosek told the *Observer*. “If they do not, the clear progress on improving the research culture will inevitably regress.”

This problematic structure is slowly changing in the wake of the replication crisis. APS and its journal editors have led many changes to publishing practices in the spirit of transparency. These include preregistration, in which scientists detail the hypotheses, designs, and analysis plans for their studies before collecting data and analyzing results. And they're awarding Open Practices badges to authors who make their data and materials openly available for examination. Evaluations of the impact of these advances are still in the early stages. Surveys have yielded variable rates of data sharing and preregistration, Nosek and colleagues reported. But there are some promising signs.

In survey results published in 2019, a group of researchers led by economist Garret Christensen found that the percentage of psychological scientists who reported sharing data or code rose from 20% in 2011 to 51% in 2017; over the same period, preregistrations climbed from 8% to 44% (Christensen et al., 2019).

Meanwhile, more and more journals are adopting policies designed to boost replicability. And universities are slowly but increasingly mentioning a preference for replicability and transparency in the work of applicants for faculty positions.

Many have argued that open science practices such as preregistration are cumbersome and slow down the pace of discovery. Others worry that an escalation in replication projects will further damage the field's reputation, not to mention the careers of certain researchers.

But reformers say innovative research will be anything but stifled if the field can embrace structural changes and stop regarding replication failures as a crisis. Science exists to expand the boundaries of knowledge, Nosek and his colleagues wrote in their *Annual Review of Psychology* article—and that expansion includes false starts and unfulfilled predictions. A healthy psychological science will produce many nonreplicable findings while continuing to strive for improved replicability, they said.

“Part of a successful scientific career,” they wrote, “involves getting used to being wrong, a lot.” ?

*Scott Sleek is a freelance writer in Silver Spring, Maryland.*

## References

Christensen, G., Wang, Z., Paluck, E. L., Swanson, N., Birke, D. J., & Littman, M. E. (2019). Open science practices are on the rise: The State of Social Science (3S) Survey. *PLOS Biology*, *14*(5), e1002460. \_

<https://doi.org/10.1371/journal.pbio.1002460>.

Carney, D. W. (2016, September 25). *My position on “Power Poses.”*

[https://faculty.haas.berkeley.edu/dana\\_carney/pdf\\_my%20position%20on%20power%20poses.pdf](https://faculty.haas.berkeley.edu/dana_carney/pdf_my%20position%20on%20power%20poses.pdf)

Ebersole, C. R., Axt, J. R., & Nosek, B. A. (2016). Scientists' reputations are based on getting it right, not being right. *PLOS Biology*, 14(5), e1002460. <https://doi.org/10.1371/journal.pbio.1002460>

Edlund, J. E., Cuccolo, K., Irgens, M. S., Wagge, J. R., & Zlokovich, M. S. (2021). Saving science through replication studies. *Perspectives on Psychological Science*. Advance online publication. <https://doi.org/10.1177/1745691620984385>

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxson, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., ... Coles, N. A. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour*, 5(1), 159–169, <https://doi.org/10.1038/s41562-020-01007-2>

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline, M., Nuijten, M. B., Rohrer, J., Romero, F., Scheel, A., Scherer, L., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*. Advance online publication. <https://doi.org/10.31234/osf.io/ksfvq>.

Oosterhof, N.N., Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, USA*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 943. <https://www.science.org/doi/10.1126/science.aac4716>

Rohrer, J. M., Tierney, W., Uhlmann, E. L., DeBruine, L. M., Heyman, T., Jones, B., Schmulke, S. C., Silberzahn, R., Willén, R. M., Carlsson, R., Lucas, R. E., Strand, J., Vazire, S., Witt, J. K., Zentall, T. R., Chabris, C. F., & Yarkoni, T. (2021). Putting the self in self-correction: Findings from the Loss-of-Confidence Project. *Perspectives on Psychological Science*. Advance online publication. <https://doi.org/10.1177/1745691620964106>.