

# Improving Social Science: Lessons from the Open Science Movement

Per Engzell, *University of Oxford and Stockholm University*

Julia M. Rohrer, *Max Planck Institute for Human Development and University of Leipzig*

Recent years have been times of turmoil for psychological science. Depending on whom you ask, the field underwent a “replication crisis” (Shrout and Rodgers 2018) or a “credibility revolution” (Vazire 2018) that might even climax in “psychology’s renaissance” (Nelson, Simmons, and Simonsohn 2018). This article asks what social scientists can learn from this story. Our take-home message is that although differences in research practices make it difficult to prescribe cures across disciplines, much still can be learned from interdisciplinary exchange. We provide nine lessons but first summarize psychology’s experience and what sets it apart from neighboring disciplines.

As a sociologist and a psychologist, we are outsiders to political science. What unites us is an interest in meta-scientific questions that has made us wonder how disciplines beyond psychology can benefit from increased transparency. Whereas we aim to address social scientists in general, our perspective is that of quantitative research. We focus on the practices of *open data*, *open materials*, and *preregistration*. These often are thought of as means to improve the credibility of research—for example, through increasing reproducibility (i.e., ensuring that a reanalysis of the same data results in the same conclusions) and/or replicability (i.e., ensuring that an empirical replication of a study leads to the same conclusions). Of course, open science also encompasses other practices such as open access publication and open educational resources, with a broad range of underlying goals, including increased accessibility and reduced inequalities.

## THE VIEW FROM PSYCHOLOGY

Psychology’s current reform movement began with the insight that certain research practices were both problematic (Simmons, Nelson, and Simonsohn 2011) and widespread (John, Loewenstein, and Prelec 2012). Low power, misuse of significance testing, researcher degrees of freedom, and post hoc hypothesizing had created a cycle in which flashy but spurious results spread with little attempt of falsification. This was exposed through a series of high-profile replication failures (e.g., Open Science Collaboration 2015) that made the

problems visible and created momentum but also caused backlash (Baumeister 2016; Gilbert et al. 2016).

The next phase was marked by attempts to solve the underlying issues through increased transparency. Journals such as the Association for Psychological Science’s flagship *Psychological Science* adopted “badges” for contributions that adhered to open standards (Lindsay 2017). By late 2018, more than 22,000 preregistrations had been filed on the Open Science Framework. More than a dozen job advertisements have asked applicants to add an open science statement to demonstrate how they have contributed to replicable, reproducible, and transparent research (see <https://osf.io/7jbnt>). Ostensibly, openness has become mainstream.

However, empirical follow-ups often have been sobering. Even with open data and open materials, analyses may be reproduced only with considerable effort or help, if at all (e.g., Hardwicke et al. 2018). Preregistrations often are too vague to keep researcher degrees of freedom at bay (Veldkamp et al. 2018), and undisclosed deviations from the preregistered plan seem to be common (Claesen et al. 2019).

We now seem to have entered a phase in which the movement’s initial success has invited a broader range of proposals not always linked to openness as such, including calls for better measurement (Flake and Fried 2019), theoretical rigor (Muthukrishna and Henrich 2019), stricter significance thresholds (Benjamin et al. 2018), and multi-model analysis (Orben and Przybylski 2019). There also is growing interest in causal inference (Rohrer 2018) and transparency in analyzing preexisting data (Weston et al. 2019)—issues long known to the political science community.

## IS PSYCHOLOGY’S EXPERIENCE GENERALIZABLE?

It may be tempting to apply some of the tools and insights from psychology to other social sciences. However, recent developments in the field have been shaped by its particularities. For example, the subfields that were hit hardest by the replication crisis stand out for their emphasis on counter-intuitive results carefully teased out in small-scale experiments. Hence, the prior probability of a tested hypothesis might be low and statistical evidence may be weak, but

empirical replication studies are comparably inexpensive—which, all else being equal, makes it easier to discover the problem.

Other social sciences place less emphasis on novelty and more on cumulative refinement of observational estimates with large-scale, representative data. Here, hypotheses may be more plausibly true to begin with and statistical evidence may be stronger, but replication on new data can be difficult or

*The statistical flukes or variance false-positives that psychology has grappled with might be overshadowed by bias false-positives from flawed sampling, measurement, or design, which can be quite replicable if follow-up studies suffer from the same flaws.*

impossible. This does not mean that these other fields are infallible but rather that problems and solutions may differ. The statistical flukes or *variance* false-positives that psychology has grappled with might be overshadowed by *bias* false-positives from flawed sampling, measurement, or design, which can be quite replicable if follow-up studies suffer from the same flaws.

Where does political science stand in all of this? Increasingly, it is a discipline that takes pride in causal inference (Clark and Golder 2015). Ironically, by moving closer to an experimental ideal, statistical flukes—that is, *variance* false-positives—become a greater concern (Young 2019). Moreover, whereas taste for novelty is arguably less of an issue than in psychology, political desirability can have similar influence (Zigerell 2017). Furthermore, certain problems that have been identified in psychology also have been pointed out in political science, including low computational reproducibility (Stockemer, Koehler, and Lentz 2018; cf. Jacoby, Lafferty-Hess, and Christian 2017) and sanitized research narratives that do not capture the actual complexity of the process (Yom 2018).

Hence, there are both commonalities and differences in the problems that affect different social sciences. With their focus

*Especially for early-career researchers, the entry barrier will be lowered as they become less dependent on access to prominent mentors and run a lower risk of wasting time on a topic known to be “doomed” by insiders.*

on increased transparency, open science practices might be able to attenuate some of them. How these practices can be implemented, however, will depend on the methods and approaches used by researchers—which vary between and within different social sciences. Indeed, political science covers a wide range of methods and approaches. Thus, the lessons we suggest are broader points on a metalevel rather than specific prescriptions.

## LESSONS FOR IMPROVING SOCIAL SCIENCE

We draw the following lessons from psychology’s experience.

### One Size Does Not Fit All

Reform attempts in psychology have had an impact precisely because they struck at some of the field’s central shortcomings. Our first lesson, therefore, is that attempts to improve the empirical status of a discipline must be localized to that

discipline. This work could begin by asking a set of basic questions: Which criteria are used to judge scientific progress, and how are scientific claims evaluated (e.g., Elman, Kapiszewski, and Lupia, 2018)? Which problems are the biggest threat to inference? What are current norms and what keeps researchers from abandoning those that are counterproductive? Once these issues have been settled and proposals are being evaluated, we must consider costs and benefits, division of labor, incentive design, and so on.

### Harness Tacit Knowledge

Where to begin then? To some extent, the prevalence of specific (mal-)practices can be surveyed empirically (John, Loewenstein, and Prelec 2012) and their impact can be gauged formally (Smaldino and McElreath 2016), as can the potential effects of proposed solutions (Smaldino, Turner, and Kallens 2019). However, the first step should be an open and critical dialogue among researchers in the field. In our experience, knowledge of bad practices can be widespread without leading to action. It is, for example, telling that experiments with prediction markets have found that scholars seem quite capable of identifying the replications least likely to succeed (Dreber et al. 2015). Such tacit know-

ledge, once it becomes explicated, is an important resource for improving science.

### Assess the Benefits of Open Science...

Why would we want transparency in the first place? For some, the ability to reproduce an analysis is the only way to fully understand and evaluate it (King 1995, 444). However, the benefits of transparency extend beyond critical evaluation.

---

---

Sharing of data and other materials reduces duplicate work and increases the yield from a given dataset, enables pooling of evidence, imposes greater self-scrutiny, and allows others to adapt and build on existing efforts. These benefits serve credibility as well as other goals including efficiency and equality. Especially for early-career researchers, the entry barrier will be lowered as they become less dependent on access to prominent mentors and run a lower risk of wasting time on a topic known to be “doomed” by insiders.

### **...As Well as the Costs, and Ways to Reduce Them**

The costs of open science are real. Considering the social costs, much of the recent backlash has been driven by targets of scrutiny who felt unfairly treated. This is an issue of culture, as Janz and Freese (2020) discuss in this symposium. Considering the practical costs, transparency requires work. An obligation to share materials can shift incentives away from original data collection or lead informants to withhold sensitive information (Connors, Krupnikov, and Ryan 2019). Some of these drawbacks have technical solutions. To preserve confidentiality, there has been experimentation with “synthetic data” that preserve joint distributions without exposing individuals (Nowok, Raab, and Dibben 2016). As for the workload of preregistrations, standard operating procedures can shorten the process (Lin and Green 2016). Moreover, push-button replications (Khatua 2018) can be used to verify analysis pipelines and markup-based tools (Hardwicke 2018) to detect undisclosed deviations between preregistration and manuscripts.

### **Beware of Tokenism**

Psychology’s mixed experience with badges highlights how—when changed rules and norms bring new incentives—there is always a temptation to cut corners. As with any target, open practices risk turning into another metric that researchers game for their own gain. To some extent, clearer standards—more transparency about transparency—might clarify what is

called out for doing openness wrong—for example, claiming that a study was preregistered despite substantial deviations in the final publication. This is far from a fair solution in which the same checks are consistently applied to everyone. However, such a fair solution seems to be necessary if open practices are to become established. There are various ways to assign that burden—ranging from editorial boards and reviewers, to universities and institutes, to students as part of the curriculum (King 2006). A real commitment to openness may require a new professional role dedicated to verification. This does not seem outlandish given the growing cadre of administrators tasked with facilitating research.

### **Reward Public-Good Contributions**

That a finding is reproducible using the same data and analysis is, admittedly, a low bar. Other forms of replication involve applying different methods to the same data or the same method to different data (Freese and Peterson 2017). Authors have few incentives to support this type of generative work because there is no good system for adequately crediting materials that help others. Someone who spent hundreds of hours gathering, cleaning, and analyzing a dataset will be reluctant to share the fruits of that labor without reward. Fair recognition of public-good contributions might counteract some of the shortcomings of gameable “checkbox” policies, such as badges and mandatory code sharing. For example, hiring committees could explicitly consider “secondary research value,” such as new insights generated on the basis of data openly shared by applicants, regardless of whether they coauthored the respective manuscript.

### **Be Inclusive**

In our view, one of the main benefits of open science is its inclusionary aspect. By widening access to information and lowering entry barriers, it promises to be both more democratic and more efficient than the status quo. However, open science also can create barriers. Power struggles are inherent

*Seen this way, the recent push toward openness is neither a fad nor an innovation but simply a recognition of our shared interest as a scientific community. This leads to an uplifting conclusion: the aims of open science are largely those of the scientific method itself—that is, open science is really just science.*

expected. Does open data merely indicate that some data have been made available, or should it also be the right data to reproduce the numbers? Does preregistered mean “there is a document that you can compare to the published article” or that the analyses reported were conducted as prespecified unless declared otherwise? These are only partial solutions, and we also must consider the division of labor and incentive structure.

### **Mind the Division of Labor**

A crucial question is: Who should check whether materials allow for reproducing findings? Right now, the answer seems to be “anybody who feels like it.” Occasionally, researchers are

in institutional change and, in science, traceable at least to the dawn of the experiment (Shapin and Schaffer 1985). From the inside, the open science movement looks generous and inspiring, but it can appear differently for those who feel left out. For example, higher standards for computational reproducibility require skills that not everyone has had an opportunity to acquire. Creating accessible resources therefore should be a central part of promoting open science. There also is a risk that open science is perceived as a cliquish movement pushed by zealots that must be actively worked against—as in any group effort, cohesion must be balanced with inclusiveness.

## Open Science Is Just Science

If open science has any unifying core, it is the shared understanding that increased transparency and accessibility can improve the quality of research and keep scientists' biases in check. We noticed that—more often than not—the desire for such improvement stems from a wish to answer meaningful research questions with real-world implications rather than an interest in transparency as an end in itself. Seen this way, the recent push toward openness is neither a fad nor an innovation but simply a recognition of our shared interest as a scientific community. This leads to an uplifting conclusion: the aims of open science are largely those of the scientific method itself—that is, open science is really just science.

## ACKNOWLEDGMENTS

Per Engzell acknowledges funding from the Swedish Research Council for Health, Working Life, and Welfare (FORTE), Grant No. 2016-07099, and support from Nuffield College and the Leverhulme Centre for Demographic Science, the Leverhulme Trust. Both authors contributed equally and are listed in alphabetical order. ■

## REFERENCES

- Baumeister, Roy F. 2016. "Charting the Future of Social Psychology on Stormy Seas: Winners, Losers, and Recommendations." *Journal of Experimental Social Psychology* 66:153–58.
- Benjamin, Daniel J., James O. Berger, Magnus Johannesson, Brian A. Nosek, E. J. Wagenmakers, Richard Berk, et al. 2018. "Redefine Statistical Significance." *Nature Human Behaviour* 2 (1): 6–10.
- Claesen, Aline, Sara Gomes, Francis Tuerlinckx, and Wolf Vanpaemel. 2019. "Preregistration: Comparing Dream to Reality." *PsyArXiv*, May 9. Available at <https://doi.org/10.31234/osf.io/d8wex>.
- Clark, William Roberts, and Matt Golder. 2015. "Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?" *PS: Political Science & Politics* 48 (1): 65–70.
- Connors, Elizabeth C., Yanna Krupnikov, and John Barry Ryan. 2019. "How Transparency Affects Survey Responses." *Public Opinion Quarterly* 83 (S1): 185–209.
- Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A. Nosek, and Magnus Johannesson. 2015. "Using Prediction Markets to Estimate the Reproducibility of Scientific Research." *Proceedings of the National Academy of Sciences* 112 (50): 15343–47.
- Elman, Colin, Diana Kapiszewski, and Arthur Lupia. 2018. "Transparent Social Inquiry: Implications for Political Science." *Annual Review of Political Science* 21:29–47.
- Flake, Jessica Kay, and Eiko I. Fried. 2019. "Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them." *PsyArXiv*, January 17. Available at <https://doi.org/10.31234/osf.io/hs7wm>.
- Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science." *Annual Review of Sociology* 43:147–65.
- Gilbert, Daniel T., Gary King, Stephen Pettigrew, and Timothy D. Wilson. 2016. "Comment on Estimating the Reproducibility of Psychological Science." *Science* 351 (6277): 1037.
- Hardwicke, Tom E. 2018. "SMART Preregistrations." Presentation at the Society for the Improvement of Psychological Science (SIPS) 2018 Meeting, Grand Rapids, MI. Available at <https://osf.io/t8yjb>.
- Hardwicke, Tom E., Maya B. Mathur, Kyle MacDonald, Gustav Nilsson, George C. Banks, Mallory C. Kidwell, Alicia Hofelich Mohr, et al. 2018. "Data Availability, Reusability, and Analytic Reproducibility: Evaluating the Impact of a Mandatory Open Data Policy at the Journal Cognition." *Royal Society Open Science* 5 (8): 180448.
- Jacoby, William G., Sophia Lafferty-Hess, and Thu-Mai Christian. 2017. "Should Journals Be Responsible for Reproducibility?" *Inside Higher Ed: Rethinking Research*, July 17. Available at [www.insidehighered.com/blogs/rethinking-research/should-journals-be-responsible-reproducibility](http://www.insidehighered.com/blogs/rethinking-research/should-journals-be-responsible-reproducibility).
- Janz, Nicole, and Jeremy Freese. 2020. "Replicate Others as You Would Like to Be Replicated Yourself." *PS: Political Science & Politics*. doi:10.1017/S1049096520000943.
- John, Leslie K., George Loewenstein, and Drazen Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science* 23 (5): 524–32.
- Khatua, Sayak. 2018. "What's the Deal with Push Button Replications?" *International Initiative for Impact Evaluation (3ie)*. Available at <http://3ieimpact.org/blogs/whats-deal-push-button-replications>.
- King, Gary. 1995. "Replication, Replication." *PS: Political Science & Politics* 28 (3): 444–52.
- King, Gary. 2006. "Publication, Publication." *PS: Political Science & Politics* 39 (1): 119–25.
- Lin, Winston, and Donald P. Green. 2016. "Standard Operating Procedures: A Safety Net for Pre-Analysis Plans." *PS: Political Science & Politics* 49 (3): 495–500.
- Lindsay, D. Stephen. 2017. "Sharing Data and Materials in Psychological Science." *Psychological Science* 28 (6): 699–702.
- Muthukrishna, Michael, and Joseph Henrich. 2019. "A Problem in Theory." *Nature Human Behaviour* 3:221–29.
- Nelson, Leif D., Joseph Simmons, and Uri Simonsohn. 2018. "Psychology's Renaissance." *Annual Review of Psychology* 69:511–34.
- Nowok, Beata, Gillian M. Raab, and Chris Dibben. 2016. "Synthpop: Bespoke Creation of Synthetic Data in R." *Journal of Statistical Software* 74 (11): 1–26.
- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716.
- Orben, Amy, and Andrew K. Przybylski. 2019. "The Association between Adolescent Well-Being and Digital Technology Use." *Nature Human Behaviour* 3:173–82.
- Rohrer, Julia M. 2018. "Thinking Clearly about Correlations and Causation: Graphical Causal Models for Observational Data." *Advances in Methods and Practices in Psychological Science* 1 (1): 27–42.
- Shapin, Steven, and Simon Schaffer. 1985. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press.
- Shrout, Patrick E., and Joseph L. Rodgers. 2018. "Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis." *Annual Review of Psychology* 69:487–510.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22 (11): 1359–66.
- Smaldino, Paul E., and Richard McElreath. 2016. "The Natural Selection of Bad Science." *Royal Society Open Science* 3 (9): 160384.
- Smaldino, Paul E., Matthew A. Turner, and Pablo Andrés Contreras Kallens. 2019. "Open Science and Modified Funding Lotteries Can Impede the Natural Selection of Bad Science." *OSF Preprints*, January 28. Available at <https://doi.org/10.1098/rsos.190194>.
- Stockemer, Daniel, Sebastian Koehler, and Tobias Lentz. 2018. "Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature." *PS: Political Science & Politics* 51 (4): 799–803.
- Vazire, Simine. 2018. "Implications of the Credibility Revolution for Productivity, Creativity, and Progress." *Perspectives on Psychological Science* 13 (4): 411–17.
- Veldkamp, Coosje L. S., Marjan Bakker, Marcel A. L. M. van Assen, Elise A. V. Cromptoets, How H. Ong, Brian A. Nosek, Courtney K. Soderberg, David Mellor, and Jelte M. Wicherts. 2018. "Ensuring the Quality and Specificity of Preregistrations." *PsyArXiv*, September 4. Available at <https://doi.org/10.31234/osf.io/cdgyh>.
- Weston, Sara J., Stuart J. Ritchie, Julia M. Rohrer, and Andrew K. Przybylski. 2019. "Recommendations for Increasing the Transparency of Analysis of Preexisting Datasets." *Advances in Methods and Practices in Psychological Science* 2 (3): 214–27.
- Yom, Sean. 2018. "Analytic Transparency, Radical Honesty, and Strategic Incentives." *PS: Political Science & Politics* 51 (2): 416–21.
- Young, Alwyn. 2019. "Consistency Without Inference: Instrumental Variables in Practical Application." London School of Economics: Unpublished Manuscript.
- Zigerell, L. J. 2017. "Reducing Political Bias in Political Science Estimates." *PS: Political Science & Politics* 50 (1): 179–83.