The Division of Replication Labor

Romero, Felipe

Published in:
Philosophy of Science

DOI:
10.1086/710625

The Division of Replication Labor

Felipe Romero*†

Scientists are becoming increasingly aware of a "replicability crisis" in the behavioral, social, and biomedical sciences. Researchers have made progress identifying statistical and methodological causes of the crisis. However, the social structure of science is also to blame. In the fields affected by the crisis, nobody is explicitly responsible and rewarded for doing confirmation and replication work. This article makes the case for a social structural reform to address the problem. I argue that we need to establish a reward system that supports a dedicated group of confirmation researchers and formulate a proposal that would achieve this.

1. Introduction. In many productive spheres in society, when we really care about a job being done right, someone other than the person who does the job has the specific task of verifying that this is the case, even if sporadically and randomly. You would not trust a food company that bypasses inspection by the Food and Drug Administration. Nor would you like to fly on an airplane that runs with software that has not been independently tested. And although nobody enjoys a tax audit, most of us agree that there should be tax auditors. Given that we value these jobs, there is a reward system for them. Should we not treat science with as much care?

Scientists are becoming increasingly aware of a "replicability crisis" in the behavioral, social, and biomedical sciences (Baker 2016). In recent years, independent researchers have unexpectedly failed to replicate many findings (i.e., when they repeat the original experiment they do not obtain the original result). The estimates of replicability success are worryingly low: 36% in experimental psychology (Open Science Collaboration 2015) and 11%–20% in cancer research (Prinz, Schlange, and Asadullah 2011; Begley and Ellis

*To contact the author, please write to: Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands; e-mail: c.f.romero@rug.nl.

†I am grateful to Ryan Doody, Jan Sprenger, Teresa Ai, Carl Craver, and the editor for their useful comments on previous drafts. I also thank the audience at the Philosophy of Science Association Biennial Meeting, Seattle, November 2018, for helpful discussion.

Philosophy of Science, 87 (December 2020) pp. 1014–1025. 0031-8248/2020/8705-0020$10.00

2012). Many of the findings that fail to replicate have not only been published in prominent journals but also influenced other scientists and the public.

Researchers have made progress identifying statistical and methodological causes of the crisis (Ioannidis 2008; Simmons, Nelson, and Simonsohn 2011; John, Loewenstein, and Prelec 2012; see Romero 2019, for review) and forcefully defended statistical and methodological reforms (Cumming 2012; Chambers 2013; Lee and Wagenmakers 2013; Mayo 2018; Machery 2019).

Now, the social structure of science is also to blame. In theory, confirmation, which includes replication among other practices, is an essential step in the scientific process (but see Leonelli 2018; Feest 2019). However, in the fields affected by the crisis, nobody is explicitly responsible and rewarded for doing confirmation work. Scientists in these fields work under a reward system that encourages novelty at the expense of careful confirmation efforts. Career pressures encourage them to rush low-quality research into print and place too much trust in unreliable research. Quality control in these fields relies on peer review, which at best establishes the plausibility of findings but never their reliability.

But, can we change the social structures of science to make science more replicable? If so, what would the suitable structures look like? These critical questions remain unaddressed. (And indeed, this is an area ripe for social epistemological work and philosophy of science in practice.) This article takes steps to address them and proposes a social structural reform. Like for food inspection, software testing, and tax audits, I argue that we should treat confirmation in science with the care that it deserves and sketch a proposal that does precisely this in the context of laboratory-based research with convenience samples. This proposal suggests establishing a reward system that supports a dedicated group of confirmation researchers. Transforming the social structure of science might seem a utopian goal. But we need to bring the social structure to the forefront of the replicability crisis discussion. Without devising alternatives to the current system, statistical and methodological reforms will fall short in addressing the crisis.

The remainder of the article is organized as follows. Section 2 explains why confirmation research is essential for scientific self-correction and how confirmation and, in particular, replication work is neglected. Section 3 presents my proposal, which I call the "professional scheme." Section 4 presents three arguments in favor of this scheme, and section 5 discusses three objections.

2. Replication: Three Problems. Philosophers and methodologists distinguish *exploratory research* and *confirmatory research* (Steinle 1997; Sakaluk 2016). The former is research that looks for patterns, often in an unguided way, while the latter tests predefined hypotheses. In practical terms, if the hypothesis of a study is formulated before the study is conducted (and not changed while it is conducted or afterward), then the research is confirmatory. Although different, these two types of research are related. The results of exploratory projects inspire hypotheses that researchers later approach in a confirmatory mode. And some projects lay somewhere in an exploratory-confirmatory continuum (Wright 2017).

Replication is an exemplary confirmatory practice. There are different definitions of the term (see Romero 2019, for discussion). For my purposes here, replication is an experiment that mirrors an original experimental design in all factors that are purportedly causally responsible for the effect. In a typical replication project, the replicator tests the same hypothesis specified in the original study using the same methods. It is by conducting this sort of experiment and failing at alarming rates that the crisis emerged. Now, while replication has been the focus of the crisis, replication does not exhaust confirmation. Other critical confirmatory practices are *reproduction* (i.e., rerunning analysis over preexisting data), *meta-analysis*, and *theory criticism*.

Why is confirmatory research necessary? Because of its connection with scientific self-correction. In theory, philosophers regard science as self-corrective: in the long run, science corrects its errors and converges on true theories (Peirce 1901/1958; Reichenbach 1938). Confirmation work is essential in this process. A single experiment can give us a correct finding, but we do not know the finding is correct without proper confirmation. The finding can be the result of a lucky accident or other sources of error. By conducting confirmation work, scientists can correct such errors. In particular, the combination of replication work and statistical inference offers one straightforward instantiation of this self-corrective process: as the number of replications of an experiment increases, the meta-analytical aggregation of their effect sizes approaches the true effect size with narrow confidence intervals.

Despite its theoretical importance, the practice of confirmation is far from the theory. The replicability crisis reveals that several subfields in the social and behavioral sciences severely neglect confirmation work and replication in particular. Specifically, we can identify three problems:

1. *Replication work is not independent.*—When a replication attempt happens, usually the original author of the finding (or close collaborators) conducts it (Makel, Plucker, and Hegarty 2012). These replications are epistemically questionable, as original authors have a conflict of interest when judging their own work. Indeed, suspiciously, replication attempts by original authors are more successful than those by independent authors (Makel et al. 2012; Kunert 2016).
2. *Replication work is not systematic.*—Very few findings are subject to replication attempts (Makel et al. 2012). Replication attempts happen in isolation (e.g., they stem from individual researchers' initiatives)

and not as a standard practice to test important findings rigorously (e.g., replicating multiple times and across different conditions.)

3. *Replication work is not sustainable*.—While many researchers acknowledge the epistemic value of replication, there are few material incentives to conduct replications (Koole and Lakens 2012; Nosek, Spies, and Motyl 2012). Relative to novel research, replication is underrewarded. Hence, researchers who want to advance in their careers are better off conducting novel research.

As these problems arise from social structural conditions, I suggest that we need social reforms to address them. I use the notion of *self-corrective labor schemes*. These schemes specify how scientists should organize their confirmation efforts, which involves establishing roles, responsibilities, and communication rules. (See Romero [2018] for a discussion of different scheme proposals.) Here I defend one scheme, which I call the *professional scheme*.

3. A Proposal: The Professional Scheme. The professional scheme proposes an alternative way of organizing confirmation work. This scheme has two main features. First, there is a specialized group of confirmation researchers. Second, there is a distinct reward system to support their work. I explain these features in turn.

3.1. Division of Labor. The first feature of the professional scheme is that a specialized group of scientists conduct confirmation work (i.e., replication, reproduction, meta-analysis, and theory criticism). Given its increasing complexity, contemporary science requires the division of cognitive labor (Kitcher 1990; Weisberg and Muldoon 2009). Currently, scientists divide their *subject matter*: they specialize and contribute to distinct fields. However, this is not the only way we can conceive the division. We can also divide cognitive labor according to *stages of the research process*. Indeed, Francis Bacon advocated for such a division in his New Atlantis novel (1627/2000), which is perhaps the earliest account of institutionalized science. In the New Atlantis, some researchers specialize in designing experiments, others conduct them, and others analyze the data and generalize findings.

In line with this idea, the professional scheme distinguishes two kinds of scientific workers. The first is the most common kind of academic scientist today, which I call *discovery researchers*. The second is a new kind, which I call *confirmation researchers*. These two kinds are distinct across three dimensions: (1) the type of research that they conduct (i.e., exploratory or confirmatory), (2) the epistemic goal that they have when approaching their research, and (3) the target findings that they study (see table 1).

On the one hand, discovery researchers engage in both exploratory and confirmatory projects—perhaps more the former, as it is the case today. Their

TABLE 1. Discovery Researchers and Confirmation Researchers

	Discovery Researchers	Confirmation Researchers
Type of research	Exploratory and confirmatory	Confirmatory
Epistemic goal	Theoretical innovation	Assess the reliability of findings
Target findings	Their own findings	Others' findings

main epistemic goal is theoretical innovation, and they work primarily in producing their own findings and sometimes analyzing findings from their colleagues. On the other hand, confirmation researchers constitute a new kind. First, the type of research that they conduct is primarily confirmatory. Second, unlike discovery researchers who prioritize innovation, the epistemic goal of confirmation researchers is to assess the reliability of findings. And third, they work with others' findings as their object of study. That is, confirmation researchers support the self-corrective process in their subfields. To be clear, the proposal is not to change how most researchers work today. Discovery researchers may still conduct confirmatory research. Instead, the proposal adds confirmation researchers for error-control purposes within their fields.

While confirmation researchers do not look for new discoveries, their work requires a high degree of skill and creativity and, therefore, should not be perceived as second-class work. Experimentally, confirmation researchers should be skilled at discovering confounding variables, evaluating the boundary conditions of effects, and optimizing experimental procedures. Analytically, they should be skilled at cutting-edge statistical tools and meta-analytic tools to evaluate large bodies of work. Additionally, since confirmation work is resource intensive, confirmation researchers should be competent at establishing and maintaining collaborations with other laboratories (e.g., conducting multisite projects). This profile should be further tailored to specific subfields. For instance, in subfields that rely heavily on secondary data analysis, confirmation researchers should be skilled at reproduction (as opposed to replication), which involves reanalysis, testing hypothesis over existing data sets, and running alternative models.

3.2. Division of Reward Systems. In the professional scheme, discovery researchers work under a *novelty-based reward system*. This is indeed the reward system that currently governs academic research. In this system, characterized by sociologists and economists of science as "the priority rule," scientists are rewarded for making new discoveries (Merton 1957; Stephan 2012). They establish priority via peer-reviewed publication, and their reward is prestige (i.e., recognition from their peers), which comes in the form of positions, career advancement, and prizes. While there are financial benefits, they are derived from the scientist's prestige and are not her primary motivation.

Now, confirmation researchers require a reward system that is not based on novelty. This is because confirmation, and replication work in particular, is not novel work. Hence, it is at odds with the priority rule (Koole and Lakens 2012; Romero 2017). Unlike discovery researchers, confirmation researchers' rewards cannot depend on being first in showing that a finding is correct.

To address this problem, the professional scheme proposes that confirmation researchers work under a *service-based reward system*. This reward system compensates them for providing their confirmation efforts and supporting the self-corrective process in their subfield. They are rewarded for conducting confirmation projects, which do not propose new hypotheses, so long as the projects are of high quality. To make the idea intuitive, think of industry research. In the industry context, there is already a thriving service-based economy for research. Biotech companies employ highly skilled staff scientists who are rewarded for executing experiments, or being technical experts as opposed to driving novel ideas or novel experimental designs. Their financial rewards are not derived from the prestige of publishing in journals and come directly from providing their services.

The implementation of a service-based reward system for academic research can be done by establishing *confirmation-research-track positions* for professors. Currently, universities allocate professors' time to different tasks (e.g., research, teaching, advising, and administration). The proposal is to create positions in which the professor's research time is exclusively allocated to confirmation work. One related precedent shows that this is plausible. Recently, principal investigators have created PhD positions for confirmation projects. A confirmation research track would extend this idea to the professorial level. To make these positions viable as a career, universities need to acknowledge confirmation work in their promotion requirements. This can be done by focusing on quality metrics of the studies (e.g., number of studies with high statistical power, the number of preregistered studies, the number of studies with open data) rather than their novelty or sheer volume. (See Schönbrodt et al. [2015] for metric examples.)

To support confirmation-research-track positions, funding agencies also need to intervene by providing steady funding for confirmation research. That is, part of the funds that they currently allocate to exploratory projects has to be consistently allocated to confirmation and replication projects. Funding agencies have already set a precedent in this respect. For instance, the Laura and John Arnold Foundation and the Netherlands Organization for Scientific Research launched pilot programs to fund replication projects (Center for Open Science 2013; NWO 2016). An effective intervention to support confirmation researchers would be to make these programs standard.

4. Arguments for the Professional Scheme. In this section, I explain the independence, systematicity, and sustainability problems in more detail and why the professional scheme offers a solution to them.

4.1. Independence Argument. Let us examine the problem that replications are not independent. In the current system, when scientists conduct replication work, they need specific outcomes from that work to further their careers. We have two scenarios. First, when scientists replicate their own work or work that supports their own theoretical commitments, they need the outcome of the replication to be successful. For if they fail, they would be shooting themselves in the foot—perhaps even contradicting years of their own work. The second scenario is when scientists replicate someone else's work. In this case, they need the replication to fail. If they succeed, they would be nothing but second stringers. Failing, however, may give them visibility for proving the original author wrong. This would be indeed the most desirable outcome for the replicator if the target finding contradicts her theoretical commitments.

Now, if scientists need specific outcomes from their replication attempts, then they have a conflict of interest that puts them in an inadequate epistemic position. That is, their expectation for a specific outcome conflicts with their aim to do good science, which can result in any outcome. This conflict of interest makes the current system unsuitable to conduct independent replications. The situation is aggravated if the pressure to obtain a reward is too high. In such cases, the replicator is more likely to introduce error, even unconsciously. She may engage in questionable research practices (QRPs; John et al. 2012), HARKing (hypothesizing after results are known; Kerr 1998) or p-hacking (Simmons et al. 2011). She may find it harder to resist confirmation bias (Ioannidis et al. 2014; Nuzzo 2015) and may even engage in fraud. This may explain the evidence that authors who replicate their own work are more likely to "succeed" than other researchers (Makel et al. 2012).

By contrast, in the professional scheme, replication work is truly independent. In the professional scheme, confirmation researchers' incentives and rewards are entirely disconnected from the outcomes of their confirmation work. They can conduct replication work without conflicting interests because they do not need their replication projects to succeed or fail in supporting a hypothesis. In particular, unlike the scientist who replicates her own work, the confirmation researcher is not invested in furthering a particular theory. And unlike the scientist who replicates other's work, the confirmation researcher does not conduct replications with the expectation of proving the other wrong.

4.2. Systematicity Argument. Many replication failures lead to epistemically justified disagreements between the original experimenter and the replicator. This is especially the case when we have only one experiment, and only one failed replication attempt, as is often common. In these cases, it is possible to question the epistemic import of either on at least four grounds: (1) random variation, (2) the commonality of QRPs (John et al. 2012), (3) undiscovered mediators and moderators (Cesario 2014), or (4) the difficulties of importing original designs from other labs (Bissell 2013).

When replications are systematic, it is feasible to overcome these disagreements, but often they are not. Systematic replication occurs when a finding is rigorously tested across factors that could introduce variation. That is, replication of the finding has been attempted multiple times by different experimenters at different laboratories with different populations. As an example, multisite replication projects that have engaged in systematic work have successfully assessed the robustness of important findings (Open Science Collaboration 2012; Klein et al. 2014; Ebersole et al. 2016). Although not all findings deserve such a rigorous treatment, very few findings get it.

The professional scheme creates appropriate conditions for systematic replication. A confirmation researcher's job is precisely to evaluate the robustness of findings rigorously. Researchers can do this in two ways. First, they can explore the boundary conditions of effects, manipulating potential mediators and moderators and uncovering the details of the mechanisms responsible for the phenomena. Second, they can coordinate large-scale multisite projects to obtain more precise estimates of parameters of interest, in particular, effect sizes. Notice that confirmation researchers are in a better position to engage in these projects than discovery researchers. These projects result in papers and reports with dozens of authors and therefore provide few incentives for the discovery researcher.

The professional scheme has further advantages that arise from combining systematic and independent replication. Imagine we have resources to conduct 10 replications of an experiment. Consider two alternative scenarios. In scenario 1, we allocate resources to one scientist to replicate the experiment 10 times and pool the results. In scenario 2, we allocate resources to 10 independent scientists to replicate the experiment one time and pool the results. Assume that sample sizes, materials, experience, and so on, are the same in both scenarios. Scenario 2 is epistemically better than scenario 1 because scenario 2 is both systematic and independent, whereas scenario 1 is only systematic. Scenario 2 is more feasible under the professional scheme than under the current system.

Another advantage of combining systematic and independent replication is sociological. Some authors these days interpret replication attempts as aggressive personal attacks (see Yong [2012] and Bohannon [2014], who document examples of this). While these reactions are extreme, they should not surprise us given that replication attempts are rare and there is a prize for replication failure. If replication becomes a systematic practice conducted by independent confirmation researchers, the debates after a replication failure can focus on epistemic aspects of the research being questioned rather than alleged questionable motives.

4.3. Sustainability Argument. As we saw, a novelty-based reward system, in general, does not reward and hence does not incentivize replication work. Also, the only scenario in which researchers are potentially rewarded (i.e., proving an original scientist wrong) creates a conflict of interest. This situation undermines the independence of replication. Additionally, under a novelty-based reward system, replicating a finding multiple times is strongly discouraged. This undermines systematic replication.

Some authors have advocated for changes in the publication system to reward replication. Already in the 1970s, scientists tried (and failed) to establish a journal for replication research (Campbell and Jackson 1979). Recently, online archives, such as the PsychFileDrawer website, have provided a venue too. And prestigious journals have also started doing their part in this respect (Simons, Holcombe, and Spellman 2014; Vazire 2015).

However, rewarding replications with a publication is insufficient to make the practice of replication standard. This is because replication work, even if publishable, is still optional work. Moreover, when competition is high, optional work is relegated. Scientists have pressure to produce novel work to keep afloat during the competition and lack an equally intense pressure to do confirmation work. This situation undermines their epistemic motivations to find the truth. In other words, researchers lack career incentives to do confirmation work.

To address this sustainability problem, the professional scheme makes confirmation work a standard practice for a subgroup of scientists. This is possible because of the separation of reward systems for discovery researchers and confirmation researchers. This separation brings several advantages. First, the separation creates clear expectations for all researchers. Second, confirmation work is no longer relegated as optional because confirmation researchers are explicitly required and rewarded to conduct it. Third, given that confirmation researchers do not have the pressure to establish novel results, they can sustain confirmation work independently and systematically.

5. Three Objections

5.1. Confirmation Work Is Uninteresting. Someone could think that confirmation work is uninteresting. Would any researcher like to do only confirmation work? I have three responses. First, the categories of *epistemically necessary research* and *interesting research* often do not overlap. If you think that confirmation work is epistemically necessary and uninteresting, then you should agree that we should establish a system that rewards the scientists who do it. Second, confirmation work as I have characterized it (i.e., work that requires a high degree of skill and involves replication, reproduction, meta-analysis, and theory criticism) is not necessarily uninteresting. Third, rewards determine what is interesting. If confirmation work is a career option, talented scientists who value science not primarily for the thrill of discovery will want to do it.

5.2. Scientific Classism. Would separating researchers into discovery researchers and confirmation researchers produce scientific classism, with discovery researchers on top? I have two responses. First, both kinds of researchers contribute to two different but equally necessary stages of the research process. Hence, there is not a justifiable hierarchy in terms of the value of their work. Second, to reflect this in practice, as I have specified, confirmation researchers should work in confirmation research tracks that offer the same career development opportunities that discovery researchers have. Universities should not create hierarchies in terms of salary and other benefits between discovery researchers and confirmation researchers. Thus, the professional scheme would not create scientific classism.

5.3. Cost Efficiency. Is the professional scheme a cost-efficient solution to the replicability crisis? My response is that I have argued that some scientists in the community should be explicitly responsible for conducting confirmation work because the status quo is far from efficient. But I have not indicated what proportion should do this job. This issue needs to be further studied. This could be done using computer simulation work and asking questions such as what distributions of agent types (i.e., discovery researchers and confirmation researchers) are optimal given a variety of epistemic goals (e.g., increasing replicability rates to a percentage goal). I leave this study for a future occasion.

6. Conclusion. The replication crisis reveals that there is a mismatch between the theory and the practice of scientific self-correction in the social and behavioral sciences. Nobody is responsible or rewarded for conducting confirmation work. To solve this problem, I have proposed repositioning confirmation in science as a professional activity. The professional scheme achieves this by (1) distinguishing discovery researchers and confirmation researchers and (2) establishing a distinct reward system for the latter. This way, we would make replication work independent, systematic, and sustainable. Intervening on the social structure of science is difficult, given that science is a decentralized system. Hence, we need a variety of interventions to establish such a system, and they require aligning multiple parties, from funding agencies to universities and departments. However, without seriously rethinking scientific institutions, the improvements that statistical and methodological reforms can bring will lack the proper platform.

REFERENCES

Bacon, F. 1627/2000. *New Atlantis*. Project Gutenberg. http://www.gutenberg.org/ebooks/2434.
Baker, M. 2016. “1,500 Scientists Lift the Lid on Reproducibility.” *Nature* 533 (7604): 452–54.
Begley, C. G., and L. M. Ellis. 2012. “Drug Development: Raise Standards for Preclinical Cancer Research.” *Nature* 483:531–33.

Bissell, M. 2013. "Reproducibility: The Risks of the Replication Drive." *Nature* 503:333–34.
Bohannon, J. 2014. "Replication Effort Provokes Praise—and 'Bullying' Charges." *Science* 344:788–89.
Campbell, K. E., and T. T. Jackson. 1979. "The Role and Need for Replication Research in Social Psychology." *Replications in Social Psychology* 1 (1): 3–14.
Center for Open Science. 2013. "Reproducibility Initiative Receives $1.3M Grant to Validate 50 Landmark Cancer Studies." Press release, October 16. https://cos.io/about/news/reproducibility-initiative-receives-grant-validate-50-landmark-cancer-studies/.
Cesario, J. 2014. "Priming, Replication, and the Hardest Science." *Perspectives on Psychological Science* 9:40–48.
Chambers, C. D. 2013. "Registered Reports: A New Publishing Initiative at *Cortex*." *Cortex* 49:609–10.
Cumming, G. 2012. *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-analysis*. New York: Routledge.
Ebersole, C. R., et al. 2016. "Many Labs 3: Evaluating Participant Pool Quality across the Academic Semester via Replication." *Journal of Experimental Social Psychology* 67:68–82.
Feest, U. 2019. "Why Replication Is Overrated." *Philosophy of Science* 86 (5): 895–905.
Ioannidis, J. P. A. 2008. "Why Most Discovered True Associations Are Inflated." *Epidemiology* 19:640–48.
Ioannidis, J. P. A., M. R. Munafò, P. Fusar-Poli, B. A. Nosek, and S. P. David. 2014. "Publication and Other Reporting Biases in Cognitive Sciences: Detection, Prevalence, and Prevention." *Trends in Cognitive Sciences* 18:235–41.
John, L. K., G. Loewenstein, and D. Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science* 23:524–32.
Kerr, N. L. 1998. "Harking: Hypothesizing after the Results Are Known." *Personality and Social Psychology Review* 2:196–217.
Kitcher, P. 1990. "The Division of Cognitive Labor." *Journal of Philosophy* 87:5–22.
Klein, R. A., et al. 2014. "Investigating Variation in Replicability: A 'Many Labs' Replication Project." *Social Psychology* 45:142–52.
Koole, S. L., and D. Lakens. 2012. "Rewarding Replications." *Perspectives on Psychological Science* 7:608–14.
Kunert, R. 2016. "Internal Conceptual Replications Do Not Increase Independent Replication Success." *Psychonomic Bulletin and Review* 23 (5): 1631–38.
Lee, M. D., and E.-J. Wagenmakers. 2013. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.
Leonelli, S. 2018. "Re-thinking Reproducibility as a Criterion for Research Quality." In *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*, 129–46. Bingley: Emerald.
Machery, E. 2019. "The Alpha War." *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-019-00440-1.
Makel, M. C., J. A. Plucker, and B. Hegarty. 2012. "Replications in Psychology Research: How Often Do They Really Occur?" *Perspectives on Psychological Science* 7:537–42.
Mayo, D. 2018. *Statistical Inference as Severe Testing: How to Get beyond the Science Wars*. Cambridge: Cambridge University Press.
Merton, R. K. 1957. "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22:635–59.
Nosek, B. A., J. R. Spies, and M. Motyl. 2012. "Scientific Utopia." Pt. 2, "Restructuring Incentives and Practices to Promote Truth over Publishability." *Perspectives on Psychological Science* 7:615–31.
Nuzzo, R. 2015. "How Scientists Fool Themselves—and How They Can Stop." *Nature* 526:182–85.
NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek). 2016. "NWO Makes 3 Million Available for Replication Studies Pilot." NWO, July 19. https://www.nwo.nl/en/news-and-events/news/2016/nwo-makes-3-million-available-for-replication-studies-pilot.html.
Open Science Collaboration. 2012. "An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science." *Perspectives on Psychological Science* 7:657–60.
———. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349: aac4716.

Peirce, C. S. 1901/1958. "The Logic of Drawing History from Ancient Documents." In *The Collected Papers of Charles Sanders Peirce*, ed. A. W. Burks, vol. 4, 89–107. Cambridge, MA: Belknap.

Prinz, F., T. Schlange, and K. Asadullah. 2011. "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?" *Nature Reviews Drug Discovery* 10: 712.

Reichenbach, H. 1938. *Experience and Prediction*. Chicago: University of Chicago Press.

Romero F. 2017. "Novelty versus Replicability: Virtues and Vices in the Reward System of Science." *Philosophy of Science* 84 (5): 1031–43.

———. 2018. "Who Should Do Replication Labor?" *Advances in Methods and Practices in Psychological Science* 1 (4): 516–37.

———. 2019. "Philosophy of Science and the Replicability Crisis." *Philosophy Compass* 14 (11): e12633.

Sakaluk, J. K. 2016. "Exploring Small, Confirming Big: An Alternative System to the New Statistics for Advancing Cumulative and Replicable Psychological Research." *Journal of Experimental Social Psychology* 66:47–54.

Schönbrodt, F., M. Heene, M. Maier, and M. Zehetleitner. 2015. "The Replication-/Credibility-Crisis in Psychology: Consequences at LMU?" Center for Open Science. https://osf.io/nptd9/.

Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22:1359–66.

Simons, D. J., A. O. Holcombe, and B. A. Spellman. 2014. "An Introduction to Registered Replication Reports at *Perspectives on Psychological Science*." *Perspectives on Psychological Science* 9:552–55.

Steinle, F. 1997. "Entering New Fields: Exploratory Uses of Experimentation." *Philosophy of Science* 64:S65–S74.

Stephan, P. E. 2012. *How Economics Shapes Science*. Cambridge, MA: Harvard University Press.

Vazire, S. 2015. Editorial. *Social Psychological and Personality Science* 7:3–7.

Weisberg, M., and R. Muldoon. 2009. "Epistemic Landscapes and the Division of Cognitive Labor." *Philosophy of Science* 76:225–52.

Wright, A. G. C. 2017. "The Current State and Future of Factor Analysis in Personality Disorder Research." *Personality Disorders: Theory, Research, and Treatment* 8:14–25.

Yong, E. 2012. "A Failed Replication Attempt Draws a Scathing Personal Attack from a Psychology Professor." Blog post. *Discover Magazine Blog*, March 10. http://blogs.discovermagazine.com/notrocketscience/2012/03/10/failed-replication-bargh-psychology-study-doyen/.