

Learning from replication

Replication studies determine both the validity of scientific conclusions and provide insights into the type of methods and reporting that are necessary for robust results.

Replicating previously published studies is necessary to determine which results are robust and can be used as foundations for future research. The importance of using direct replication as a method to assess the evidence presented in support of a claim has become apparent in many disciplines, due, in part, to several recent studies showing high rates of replication failure. While these replication failures have been called a ‘crisis’ by many and have unfortunately cast doubt on the scientific method, this ‘crisis’ can also be seen as an opportunity to improve scientific methods of investigation and analysis across disciplines to ensure that the work that is published is robust and reliable.

In this issue, [Camerer et al.](#) report the results of the Social Sciences Replication Project, which performed direct replications of 21 experimental studies in the social sciences published in the journals *Science* and *Nature* between 2010 and 2015. Two previous replication projects, the Reproducibility Project: Psychology¹ and the Experimental Economics Replication Project², successfully replicated 36% and 61% of their target studies, respectively. Although those projects had high statistical power, they may not have been high powered enough and may have underestimated replication rates: perhaps more studies would replicate with even larger sample sizes, as these larger samples would be able to capture smaller effects than those originally reported. Despite increasing power substantially — sample sizes were, on average, approximately five times higher than the original studies — [Camerer et al.](#) found that only 62% of the studies show an effect in the same direction as the original studies. This effect, however, varies between 57% and 67%, depending on the ways in which successful replication is defined. [Camerer et al.](#) also found that effect sizes were, on average, approximately half the

value of the original effect sizes. According to the authors, these results suggest that the original studies likely contained both false positives and inflated effect sizes. In the Correspondences linked to the replication study, some of the authors of the original studies offer their perspective on the findings and discuss potential causes of the discrepancy between the results of [Camerer et al.](#) and their own.

What can we make of these results? In an accompanying News & Views, [Macleod](#) argues that there is immense potential to improve science using the details reported in this replication study and the ensuing dialogue and research they spark. This is a view we fully share. In addition to considering what replication success or failure means for the group of studies [Camerer et al.](#) examined, we encourage researchers to consider the opportunities for future research they open. These include not only experimentally investigating the potential causes for success or failure to replicate, but also taking a step back from each study to determine the common variables that differentiate those studies that replicate from those that do not. The more replication studies that are done, the more data there will be until there is enough for researchers to begin to use formal methods (such as unsupervised learning) to identify those characteristics of a study that contribute significantly to its replicability.

Replication studies such as the one reported in our pages by [Camerer et al.](#), the outputs of the burgeoning field of meta-science, as well as the collective efforts of research communities, funders, journals and other stakeholders over the past decade have already started to bear fruit. For example, preregistration is now much more common, so that researchers are not tempted to change their hypotheses to fit the data or to carry out multiple analyses until they obtain positive results. It is also becoming

common to use power analyses to pre-determine sample size, and to understand that results may not generalize across ages and populations unless the sample is representative of those ages and populations.

What can we, as a journal, do to promote replicability and to support those scientists undertaking replication studies? We actively encourage the submission of high-value replication studies and evaluate scientific advances not only by their conceptual or methodological novelty but also by the rigour and scale of a study, which may serve to substantially strengthen (or convincingly dispel) confidence in a scientific finding (<https://www.nature.com/nathumbehav/info/editorial-process>). Replication studies are a paradigmatic example of such advances in evidence. We encourage all researchers planning a replication study to adopt the registered report format, which allows us to accept manuscripts — in principle — for publication before the data have been collected. Although we welcome replication studies in any format, registered reports go through a two-stage peer review process (before and after data collection) to ensure that the methods and analyses are rigorous before effort and resources are invested in data collection³. There is still a long way to go, but by publishing replication studies and encouraging a broad conversation about their outcomes, we hope that we will help propagate and foster a community that will use these failures of replication as opportunities to better hone the scientific method. □

Published online: 7 September 2018
<https://doi.org/10.1038/s41562-018-0441-1>

References

1. Open Science Collaboration *Science* **349**, aac4716 (2015).
2. [Camerer, C. F. et al.](#) *Science* **351**, 1433–1436 (2016).
3. *Nat. Hum. Behav.* **1**, 0034 (2017).