

Testing the reproducibility of social science research

August 27 2018



Credit: Charles Rondeau/public domain

Today, in *Nature Human Behavior*, a collaborative team of five laboratories published the results of 21 high-powered replications of social science experiments originally published in *Science* and *Nature*, two of the most prestigious journals in science. They failed to replicate the results of more than a third of the studies and turned up significantly

weaker evidence for the remainder compared to the original studies.

In addition, prior to conducting the replications, the team set up prediction markets for other researchers to bet money on whether they thought each of the findings would replicate. The markets were highly accurate in predicting which studies would later succeed or fail to replicate.

"It is possible that errors in the [replication](#) or differences between the original and [replication studies](#) are responsible for some failures to replicate," says Gideon Nave, an assistant professor of marketing of the University of Pennsylvania's Wharton School of Business and one of the project's leads, "but the fact that the markets predicted replication success and failure accurately in advance reduces the plausibility of these explanations."

The team included researchers from Penn, the University of Innsbruck, the Stockholm School of Economics, the New Zealand Institute for Advanced Study, the Center for Open Science, the National University of Singapore, the University of Virginia, California Institute of Technology, the University of Göteborg, Harvard University, Spotify Sweden, LMU Munich, the University of Amsterdam, and the Harbin Institute of Technology.

The researchers tried to replicate one main finding from every experimental social science paper published between 2010 and 2015 that met the team's requirements of involving randomized controlled experiments conducted either with college students or online. To extend and improve on prior replication efforts, the team obtained the original materials and received the review and endorsement of the protocols from almost all of the original authors before conducting the studies. The studies were preregistered to publicly declare the design and analysis plan, and the study design included large sample sizes so that the

replications would be likely to detect support for the findings even if they were as little as half the size of the original result.

"To ensure high statistical power," says Felix Holzmeister of the University of Innsbruck, another of the project's leaders, "the average sample size of the replication studies was about five times larger than the average sample size of the original studies."

The team found that 13 of the 21 replications, or 62 percent, showed significant evidence consistent with the original hypothesis, and other methods of evaluating replication success indicated similar results, ranging from 57 to 67 percent. Also, on average, the replication studies showed effect sizes that were about 50 percent smaller than the original studies. Together this suggests that reproducibility is imperfect even among studies published in the most prestigious journals in science.

"These results show that 'statistically significant' scientific findings," says Magnus Johannesson of the Stockholm School of Economics, another project leader, "need to be interpreted very cautiously until they have been replicated even if published in the most prestigious journals."

The prediction markets the research team established correctly predicted the outcomes for 18 of the 21 replications. Market beliefs about replication were highly correlated with replication effect sizes.

"The findings of the prediction markets suggest that researchers have advance knowledge about the likelihood that some findings will replicate," notes Thomas Pfeiffer of the New Zealand Institute for Advanced Study, another of the project leaders. The apparent robustness of this phenomenon suggests that prediction markets could be used to help prioritize replication efforts for those studies that have highly important findings but relatively uncertain or weak likelihood of replication success.

"Using [prediction markets](#) could be another way for the scientific community to use resources more efficiently and accelerate discovery," adds Anna Dreber of the Stockholm School of Economics, another project leader.

This study provides additional evidence of the challenges in reproducing published results, and addresses some of the potential criticisms of prior replication attempts. For example, it is possible that higher-profile results would be more reproducible because of high standards and the prestige of the publication outlet. This study selected papers from the most prestigious journals in science.

Likewise, a critique of the Reproducibility Project in Psychology suggested that higher-powered research designs and fidelity to the original studies would result in high reproducibility. This study had very high-powered tests, original materials for all but one study, and the endorsement of protocols for all but two studies, and yet still failed to replicate some findings and found a substantially smaller effect sizes in the replications.

"This shows that increasing power substantially is not sufficient to reproduce all published findings," says Lily Hummer of the Center for Open Science, one of the co-authors.

That there were replication failures does not mean that those original findings are false. Nevertheless, some original authors provided commentaries with potential reasons for failures to replicate. These productive ideas are worth testing in future research to determine whether the original findings can be reproduced under some conditions.

The replications undertaken in this work follow emerging best practices for improving the rigor and reproducibility of research. "In this project, we led by example, involving a global team of researchers, ," says Teck-

Hua Ho of the National University of Singapore, another project lead. "The team followed the highest standards of rigor and transparency to test the reproducibility and robustness of studies in our field."

All of the studies were preregistered on [OSF](#) to eliminate reporting bias and to commit to the design and analysis plan. Also, all project data and materials are publicly accessible with the OSF registrations to facilitate the review and reproduction of the replication studies themselves.

Brian Nosek, executive director of the Center for Open Science, professor at the University of Virginia, and one of the co-authors, notes, "Someone observing these failures to replicate might conclude that science is going in the wrong direction. In fact, science's greatest strength is its constant self-scrutiny to identify and correct problems and increase the pace of discovery."

This large-scale replication project is just one part of an ongoing reformation of research practices. Researchers, funders, journals, and societies are changing policies and practices to nudge the research culture toward greater openness, rigor, and reproducibility. Nosek concludes, "With these reforms, we should be able to increase the speed of finding cures, solutions, and new knowledge. Of course, like everything else in science, we have to test whether the reforms actually deliver on that promise. If they don't, then [science](#) will try something else to keep improving."

More information: Colin F. Camerer et al, Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015, *Nature Human Behaviour* (2018). [DOI: 10.1038/s41562-018-0399-z](https://doi.org/10.1038/s41562-018-0399-z)

Provided by University of Pennsylvania

Citation: Testing the reproducibility of social science research (2018, August 27) retrieved 23 April 2024 from <https://phys.org/news/2018-08-social-science.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.