

# Comment



ILLUSTRATIONS BY DAVID PARKINS

## Replication games: how to make reproducibility research more systematic

Abel Brodeur, Anna Dreber, Fernando Hoces de la Guardia & Edward Miguel

In some areas of social science, around half of studies can't be replicated. A new test-fast, fail-fast initiative aims to show what research is hot – and what's not.

In October last year, one of us (A.B.) decided to run an ad hoc workshop at a research centre in Oslo, to try to replicate papers from economics journals. Instead of the handful of locals who were expected to attend, 70 people from across Europe signed up. The message was clear: researchers want to replicate studies.

Replication is sorely needed. In areas of the social sciences, such as economics, philosophy and psychology, some studies suggest that between 35% and 70% of published results cannot be replicated when tested with

new data<sup>1-4</sup>. Often, researchers cannot even reproduce results when using the same data and code as the original paper, because key information is missing.

Yet most journals will not publish a replication unless it refutes an impactful paper. In economics, less than 1% of papers published in the top 50 journals between 2010 and 2020 were some type of replication<sup>5</sup>. That suggests that many studies with errors are going undetected.

After the Oslo workshop, we decided to try to make replication efforts in our fields of economics and political science more systematic.

Our virtual, non-profit organization, the Institute for Replication, now holds one-day workshops – called replication games – to validate studies.

Since October 2022, we've hosted 12 workshops across Europe, North America and Australia, with 3 more scheduled this year. Each workshop has typically involved around 65 researchers in teams of 3–5 people, re-analysing about 15 papers. The teams either try to replicate papers, by generating new data and testing hypotheses afresh, or attempt to reproduce them, by testing whether the results hold if the published data are re-analysed. For many papers in our fields of study, in which the reproduction of results often involves re-running computer codes, it's possible to do much of this work in a single day (see 'A typical replication games project'). Each team's findings are released as a preprint report, and these reports will be collated and published each year as a meta-paper.

In just a few months, participants in our replication games have found papers that contain major coding errors and identified many studies that cannot be completely reproduced or replicated (along with many results that are robust). We hope to create a publicly available database of our findings later in the year.

More organized reproducibility and replication efforts similar to ours are needed. Other fields might need different formats – the papers we currently assess are mostly non-experimental, whereas re-doing experiments can take months and require specialist equipment. Yet, we think that three lessons from our experiences could help others who want to expand replication efforts.

### **Form partnerships to help scale up replication**

To assess large numbers of papers, collaborating with research centres and universities is essential. For example, our current goal is to reproduce and replicate studies in journals that have a high impact factor – specifically, 25% of empirical studies published from 2022 onwards in 8 leading economics journals and 3 leading political science journals, totalling about 350 papers per year. Then we plan to expand into other areas of the social sciences.

Having an institution that hosts our games helps us to enlist replicators. Without them, we would struggle to find specialists beyond our own fields. Having support from a university helps to raise awareness about the need for replication among local researchers, and those networks of researchers provide further opportunities to scale up replication efforts.

For instance, thanks to the contacts that were made at the games, we hope to host workshops in Kenya and Japan soon.

Broader partnerships can expand replication efforts beyond academic papers. Earlier this year, we were invited to run replication games with the International Monetary Fund (IMF) and the World Bank, to assess economics and finance papers from the two organizations. We aim to keep running these games annually, validating not only scholarly studies but also policy-oriented reports.

Establishing these relationships need not be time consuming. We've found that simply tweeting about our project and speaking about it at conferences can garner interest. That, along with word of mouth after the Oslo workshop, has been sufficient to make our project well known among economists. As a result, all the organizations that we partnered with originally contacted us – rather than the other way round – asking to get involved.

### **"It is crucial to protect researchers who check papers from career threats."**

Other researchers following in our footsteps should be aware that care is needed to avoid conflicts of interest. We receive no money from the collaborations we're involved in, because taking payment could be viewed as unethical. At the IMF and World Bank games – where people were reproducing and replicating the work of co-workers – we decided to randomly assign participants to a study, allowed them to remain anonymous and prevented participants from assessing studies authored by direct supervisors or friends.

### **Use a mediator to protect replicators**

It is crucial to protect researchers who check papers from career threats – particularly when an effort uncovers major errors. We recommend that an organization or institute mediates communication between the original study's authors and the replicators, allowing the latter to remain anonymous if they wish. One of us, acting as a representative for the Institute for Replication, serves in this capacity after each replication game.

We know that receiving an e-mail to say that someone is checking your work can be stressful. So we contact the original authors only after replicators have written up their reports, to

## **A typical replication games project**

**For each paper assessed during a workshop, researchers in each team must decide the type of validation they will perform, and plan the best way to use their time.**

Imagine a (hypothetical) paper that investigates the effect of a US government policy implemented in Texas in 2020, which uses neighbouring states as a comparison group. The paper uses data collected between 2018 and 2022 as part of a publicly available US-wide survey. The researchers conducting the study make decisions about which statistical tests to perform, which control variables to use and more.

In the replication games, a team might attempt to reproduce this paper by re-running the same analysis while making their own decisions about appropriate methods, control groups and so on.

The team might attempt to replicate it by asking whether the result remains the same if more states are used in the comparison group, or if the date range is extended to between 2015 and 2023, using data from the same survey. Or the researchers might use another survey that provides similar data to perform their replication.

Teams are formed one month before the games. Replicators read the paper and develop a plan that will allow them to do the bulk of the work on the day of the games. After the event, the replicators complete any leftover work and write a short report on their findings, which is shared with the original authors for their comments before the findings are made public.

avoid causing researchers undue worry while they wait for an effort's results. Rather than treating the discovery of errors as a 'gotcha' moment, which can put authors on the defensive, we acknowledge in our correspondence that all researchers make mistakes. To help make the process collegial, we allow authors to suggest edits to the report, and ask replicators to suggest changes to the authors' responses.

If a paper's original author won't respond to



our e-mails after several attempts over weeks or months, we still publish our replication report. This approach has led some people to argue that we should instead take our findings to the journal in which the work was published. We counter this with the possibility that some journal editors might have a conflict of interest, because publishing a retraction or correction could harm their reputation. Despite this concern, we do encourage replicators to submit their findings to the journal – but publishing a replication report as a preprint first means that other researchers can assess for themselves whether and how our findings affect the results of a paper.

So far, more than 95% of authors have answered our e-mails. In fact, many authors appreciate mediation. Often it is the first time someone has reproduced or replicated their work, and they value support and guidance in handling any errors that are uncovered. As replicating studies becomes more common, we hope that the open, professional and respectful dialogue fostered by mediation will become the norm between replicators and authors.

### Give replication personal and professional value

Busy scientists need incentives to perform replication studies. We think that having fun is the key to the replication games' success.

Many participants enjoy being involved in a progressive scholarly movement. The option of virtual participation means that researchers can take part for free, minimizing the barriers for attendance. Bringing researchers of all career stages together during our games means that junior scientists can receive mentorship, and senior researchers get the chance to brush up on practical skills such as coding, at which younger peers often excel.

**“To alleviate the fears of early-career researchers, participants should be allowed to remain anonymous.”**

Replication efforts should also offer professional incentives. Meta-papers can be extremely well-cited<sup>6</sup>. To alleviate the fears of early-career researchers, participants should be allowed to remain anonymous, with their name on the meta-paper, but not attached to a particular replication effort.

For those organizing systematic replication, meta-papers do not need to be time-consuming to generate. Our ongoing meta-papers involve taking reports from each team, which are filled in using a

template form, and compiling them into a database.

The goodwill and intellectual curiosity of scientists is sufficient to allow us to assess many papers. But we would like to broaden our scope to research that requires access to non-public data sets (administrative records, for example), and to papers involving surveys that require replicators to dedicate weeks or months to generating data and that might mean paying participants to take part. In other fields, papers involving experimental data require replicators to have the right laboratory set-up. Funders must begin to support the reproduction and replications of these types of study.

Making data and code readily available is also crucial. Our participants often spend hours trawling through data to find the variables they need, because data points are poorly labelled or defined. And papers often report only the results of analyses, rather than the raw data that were fed into them. Many journals don't require researchers to release data, and those that do often don't enforce their policies. Journals should take care to advertise and enforce editorial guidelines.

We think that efforts such as ours that normalize replication will ultimately put pressure on funders and journals to play their part. We are excited to see replication efforts in our fields – and others – continue to expand. Systematic replication has the potential to make correcting science faster. Let the games begin.

### The authors

**Abel Brodeur** is an associate professor of economics at the University of Ottawa, Canada, and is the chair at the Institute for Replication. **Anna Dreber** is a professor of economics at the Stockholm School of Economics, Sweden, and the University of Innsbruck, Austria, and is a co-director at the Institute for Replication. **Fernando Hoces de la Guardia** is an assistant project scientist with the Berkeley Initiative for Transparency in the Social Sciences at the University of California, Berkeley, California, and a co-director at the Institute for Replication. **Edward Miguel** is a professor of economics at the University of California, Berkeley, and a co-director at the Institute for Replication.  
e-mail: abrodeur@uottawa.ca

1. Open Science Collaboration. *Science* **349**, aac4716 (2015).
2. Camerer, C. F. et al. *Nature Hum. Behav.* **2**, 637–644 (2018).
3. Camerer, C. F. et al. *Science* **351**, 1433–1436 (2016).
4. Cova, F. et al. *Rev. Phil. Psychol.* **12**, 9–44 (2021).
5. Ankel-Peters, J., Fiala, N. & Neubauer, F. J. *Econ. Behav. Organ.* **212**, 219–232 (2023).
6. Botvinik-Nezer, R. et al. *Nature* **582**, 84–88 (2020).

The authors declare no competing interests.