

# Promoting Reproducibility and Replicability in Political Science

Research and Politics  
January–March 2024: 1–8  
© The Author(s) 2024  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/20531680241233439  
[journals.sagepub.com/home/rap](https://journals.sagepub.com/home/rap)  
 Sage

Abel Brodeur<sup>1</sup> , Kevin Esterling<sup>2</sup> , Jörg Ankel-Peters<sup>3</sup>,  
Natália S. Bueno<sup>4</sup>, Scott Desposato<sup>5</sup>, Anna Dreber<sup>6</sup>,  
Federica Genovese<sup>7</sup>, Donald P. Green<sup>8</sup>, Matthew Hepplewhite<sup>7</sup> ,  
Fernando Hoces de la Guardia<sup>9</sup>, Magnus Johannesson<sup>6</sup> ,  
Andreas Kotsadam<sup>10</sup>, Edward Miguel<sup>11</sup>, Yamil R. Velez<sup>8</sup> and  
Lauren Young<sup>12</sup>

## Abstract

This article reviews and summarizes current reproduction and replication practices in political science. We first provide definitions for reproducibility and replicability. We then review data availability policies for 28 leading political science journals and present the results from a survey of editors about their willingness to publish comments and replications. We discuss new initiatives that seek to promote and generate high-quality reproductions and replications. Finally, we make the case for standards and practices that may help increase data availability, reproducibility, and replicability in political science.

## Keywords

Reproducibility, replicability, political science

Ensuring the credibility of research findings hinges on the crucial role of reproduction and replication. By testing and verifying published research, both reproducibility and replication initiatives play vital roles in shaping scientific knowledge. These efforts enable us to evaluate the robustness of findings, transforming science into a self-correcting system that identifies and rectifies inaccuracies, ultimately influencing policy-making in significant ways. *Reproducibility* here is defined as testing if results and conclusions of original studies can be reproduced using the original studies' data, while *replicability* is defined as testing if results and conclusions of original studies can be repeated using new data.

When attempting to reproduce published results, researchers often face roadblocks (Colliard, 2021). Published assessment studies generally report reproducibility rates below 50%, and sometimes the success rate is single-digit (Avelino et al., 2021; Gertler Galiani and Romero, 2018; McCullough et al., 2006). This may be due to the data not

<sup>1</sup>Department of Economics, University of Ottawa and Institute for Replication, Canada

<sup>2</sup>Public Policy and Political Science, UC Riverside, USA

<sup>3</sup>RWI – Leibniz Institute for Economic Research, Germany

<sup>4</sup>Department of Political Science, Emory University, USA

<sup>5</sup>Department of Political Science, UC San Diego, USA

<sup>6</sup>Department of Economics, Stockholm School of Economics, Sweden

<sup>7</sup>Department of Politics and International Relations, University of Oxford, UK

<sup>8</sup>Political Science, Columbia University, USA

<sup>9</sup>Berkeley Initiative for Transparency in the Social Sciences, USA

<sup>10</sup>Ragnar Frisch Centre for Economic Research, Norway

<sup>11</sup>Department of Economics, UC Berkeley, USA

<sup>12</sup>Political Science, UC Davis, USA

## Corresponding author:

Abel Brodeur, Department of Economics, University of Ottawa and Institute for Replication, 120 University private, Ottawa, Ontario K1N 6N5, Canada.

Email: [abrodeur@uottawa.ca](mailto:abrodeur@uottawa.ca)



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

being publicly available because of their nature: administrative, proprietary, and/or copyrighted data (Christensen and Miguel 2018). Furthermore, for many other studies, the required computer code is not available or incomplete (Chang and Li, 2022; Gertler et al., 2018).

A few large-scale replication projects have taken place recently, including one in psychology (Open Science Collaboration, 2015), one in experimental economics (Camerer, 2016) and a social science replication project (Camerer, 2018). Replication here means that the original study's main significant result was the focus of a new study using similar methods and tests on a fresh sample. Pooling the results of these large replication projects yielded a replication rate of about 50%.

Beyond lab experiments and especially for studies based on observational data, large-scale reproduction and replication projects have not been attempted. Instead, most reproductions and replications involve reproducing or replicating the claims of one original study, often evoking lengthy debates about the interpretation of results. Yet, some recent reviews point at systematic problems such as p-hacking in studies based on observational data and that these problems are worse than for experimental/RCT data (see e.g., Brodeur et al., 2020; Young, 2017), which, if true, would translate into even lower replicability rates than for experimental studies. Experimental studies, in turn, are often said to be more prone to external and construct validity concerns, with obvious implications for their replicability in new settings (Esterling, 2023; Findley, 2021; Peters et al., 2018).

Low reproducibility and replicability rates may be due to many factors. First, many previous studies have been performed on small sample sizes or look at small effects, implying low statistical power (Ioannidis et al., 2017). Arel-Bundock (2022) assess statistical power for about 2000 articles in political science. They report that the median analysis has about 10% power. Second, there are typically many ways of testing a hypothesis, giving researchers many “researcher degrees of freedom” in their analysis (Simmons et al., 2011). Specification searching (or “p-hacking”) has been found to be a problem in political science and related disciplines (Brodeur, 2016; Gerber and Malhotra, 2008). Third, researchers might be tempted to select their hypotheses after the results are known (called “HARKing”) on the basis of whether they yield significant results (Kerr, 1998). All these factors make it hard to disentangle true results from false positive and false negative ones.

Reproduction and replication studies are an important part of changing incentives and improving the quality and credibility of original research. They may themselves also generate new knowledge and findings. Other strategies to enhance transparency and credibility in scientific research include pre-registration of hypotheses and data analysis

plans. Reproduction and pre-registration are complements, not substitutes. Pre-analysis plans can address the problem of p-hacking datasets and generating false positive results. Reproduction and replication, the focus of this paper, address the integrity and robustness of data and findings, and may allow for collecting additional data and testing new hypotheses. Recent efforts have sought to combine replication and pre-analysis plans, as exemplified by initiatives like the Metaketa initiative (<https://egap.org/our-work/the-metaketa-initiative/>), where research teams coordinate their efforts, pre-register their analyses, and often replicate the core treatments and outcomes in different settings.

In addition to the technical and logistical hurdles that prevent researchers from reproducing past evidence, the current publication incentives remain unfavorable to reproductions (Coffman et al., 2017; Clemens, 2017). Publication outlets may tend to favor novel conceptual insights over new tests of a published idea, regardless of what these tests find. Furthermore, it is possible that researchers aiming to publish reproductions as standalone projects may face incentives to engage in selective reporting, implying that reproduction efforts might also suffer from p-hacking and other questionable research practices (QRPs, see Bryan et al., 2019).

In this article, we first provide definitions for reproducibility and replicability. Next, we review data availability journal policies. We then present the results from a survey of editors of leading political science journals about their willingness to publish comments and replications. We discuss new initiatives that seek to promote and generate high-quality reproductions and replications. Last, we make the case for standards and practices that may help increase data availability, reproducibility, and replicability in political science.

## Definitions of reproducibility and replicability in political science

Several definitions of reproducibility and replicability have been used and proposed (see, e.g., Clemens 2017; Christensen and Miguel 2018; Ankel-Peters et al., 2023a). Dreber and Johannesson (2023) recently proposed definitions and indicators for economics that we summarize here and believe can be useful also for political science. As mentioned above, *reproducibility* is defined as using the original studies' data, while *replicability* is defined as using data other than what was used in the original studies.

Reproducibility is furthermore divided into three types. *Computational reproducibility* tests the extent to which results in original studies can be reproduced using the data and code from the original studies. *Recreate reproducibility* tests to what extent results in original studies can be reproduced using the information in the original studies

without access to the processed data set and/or the analysis code, while *robustness reproducibility* tests to what extent results in original studies are robust to alternative plausible analytical decisions using the same data as in the original studies.

Replicability is divided into two types. *Direct replicability* tests to what extent results in original studies can be repeated on new data using the original studies' research design and analysis (with three further sub-categories depending on if data are from the same population, a similar population, or a different population is used). *Conceptual replicability* also uses new data to test to what extent results in original studies can be repeated, but for this type of replication an alternative research design and/or alternative analysis is used to test the same hypothesis as in the original study (and conceptual replicability is also further subdivided into the same three sub-categories as direct replicability). See Dreber and Johannesson (2023) for further details and proposed indicators of reproducibility and replicability to be reported for each type.

## Lack of reproducibility and replicability in political science

Reproducibility and replication efforts contribute in essential ways to the production of scientific knowledge. Within the social sciences, political scientists have pushed the frontier of research transparency on several dimensions, such as raising the issue (King, 1995), developing guidance to rigorously document research designs (Blair, 2019), and being early adopters of data and code availability for replication (see, for instance, Bueno de Mesquita, 2003). Other contributions include developing innovative methodologies to combat p-hacking (Brezna, 2022; Young and Holsteen, 2017), proposing standard operating procedures to address omissions or ambiguities in pre-analysis plans (Lin and Green, 2016) and establishing a trusted repository to archive time-stamped registrations (EGAP).

Despite the importance of reproductions and replications for the production of scientific knowledge, progress has been slow. Existing reviews of published reproduction activities mostly document small or even miniscule replication rates (Mueller-Langer et al., 2019). The present situation is unsurprising in light of the many barriers that prevent researchers from assessing the reliability of existing research. Indeed, access to data, codes, and protocols is to date not universal in political science (Dafoe, 2014).

**Table 1** We investigate the prevalence of data availability journal policy using a sample of 28 leading political science journals (see Appendix for more details). The analysis was conducted in early June 2023. We see this analysis as an update to the review by Gleditsch and Janz (2016). We also investigated whether the journal has a data editor or

reproducibility analyst. While data editors neither check the robustness of numerical results nor their replicability, they ensure that the data and codes necessary for a computational reproducibility are available. For some journals, data editors also conduct computational reproducibility. In the event that the data cannot be shared for confidentiality reasons, the data editors typically require the authors to provide detailed information describing how other researchers may obtain the data.

With regard to the second question, four of the journals have dedicated data/replication editors, namely, Journal of Politics, Political Analysis, Political Communication, Political Science Research and Methods, and Quarterly Journal of Political Science. The American Journal of Political Science also has a verification process carried out by the Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill. This is an independent verification process that computationally reproduces numerical results for accepted articles. With regard to the first question, just one of the journals has no information on its website with regard to data or codes—Comparative Politics. Of the remaining 27 journals that were sampled, six encourage the sharing of data/codes and 21 mandate it.

It is worth emphasizing that some journals also recommend or mandate reporting standards. For instance, the *Journal of Experimental Political Science* recommends adherence to reporting standards, with a reporting guidelines checklist and guidance on how to not p-hack. *Public Opinion Quarterly* requires authors to comply with disclosure requirements in the AAPOR Code of Professional Ethics—see section III at <https://aapor.org/standards-and-ethics/>.

## Survey of editors

To examine the demand for replications among journals, we surveyed editors of leading journals in Political Science about their journals' policies to publish replications and comments (henceforth “replications,” see Figure 1 for the exact wording used in the survey). The editors were approached by email by late May 2023 and a reminder was sent early June for those who did not initially respond. The journals were selected through a crowdsourcing procedure—asking the political science Institute for Replication (I4R) board members to nominate journals and review the list. In total, 19 of 28 contacted editors responded. Figure 1 summarizes the results; the responses disaggregated by journal can be found in Table A1 in the Appendix. Most editors, 63%, stated that their journal would generally publish reproductions/replications of papers originally published in their own journal. Of those, 47% responded that they would also consider replications of papers in other journals (although for both questions, some

**Table 1.** Data availability policy and data editor.

Journal	Data availability policy?	Data editor?
American journal of political science	Yes	Independent
American political science review	Yes	No
American politics research	Encourage	No
British journal of political science	Yes	No
Comparative political studies	Yes	No
Comparative politics	No	No
European journal of political research	Yes	No
International organization	Yes	No
International studies quarterly	Yes	No
Journal of conflict resolution	Yes	No
Journal of experimental political science	Yes	No
Journal of law, economics, and organization	Yes	No
Journal of politics	Yes	Yes
J. Of public administration research & theory	Yes	No
Journal of theoretical politics	Encourage	No
Legislative studies quarterly	Encourage	No
Political analysis	Yes	Yes
Political behavior	Yes	No
Political communication	Yes	Yes
Political geography	Encourage	No
Political psychology	Encourage	No
Political research quarterly	Yes	No
Political science research and methods	Yes	Yes
Public choice	Encourage	No
Public opinion quarterly	Yes	No
Quarterly journal of political science	Yes	Yes
Research & politics	Yes	No
Review of international organizations	Yes	No

constrained this by further criteria like the replicated paper being very relevant to the journal's readership; see the detailed responses on I4R's website—<https://i4replication.org/publishing.html>). In addition, we checked the websites of these 28 journals for whether their Aims & Scope or Guide for Authors state that replications or comments are considered for publication: nine of 28 journals do so. See [Table A1](#) for journal specific details.)

Moreover, it is noteworthy that recently new opportunities to publish replications have emerged. Several journals, some new and some established, now prominently invite submissions of replications in their Aims & Scope (e.g., *Journal of Experimental Political Science* and *Political Science Research and Methods*). *Research & Politics* has replications as a category for submission:

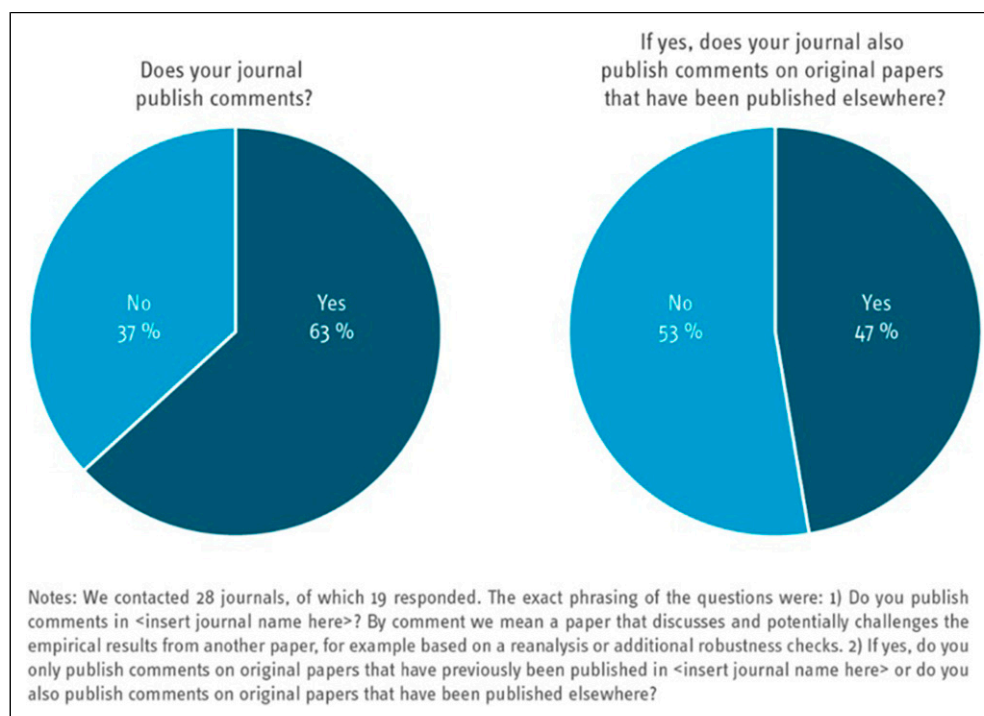
Research & Politics invites authors to consider submitting a paper that is along the lines of one or more of the following replication types:

Theoretical replication: The submitted article argues that the original theoretical model is missing at least one key element. The missing element(s) are addressed and included in the empirical analysis.

Technical replication: The submitted article identifies faults in the original research design or analysis, thereby arguing that the original results might not hold; and/or

Concept replication: The submitted article questions the validity of the original study. An alternative measurement or operationalisation is proposed which yields different substantive results.

Although taking these steps represents progress, practical challenges hinder the widespread publication of replications. These obstacles include entrenched biases favoring the status quo and difficulties in securing reviewers willing to assess debates laden with intricate technical details, demanding substantial effort. Additionally, previous efforts by journals to signal a need for replications have not



**Figure 1.** Survey among editors of leading Political Science journals. Notes: The exact phrasing of the questions were: (1) Do you publish comments in <insert journal name here>? By comment we mean a paper that discusses and potentially challenges the empirical results from another paper, for example, based on a reanalysis or additional robustness checks. (2) If yes, do you only publish comments on original papers that have previously been published in <insert journal name here> or do you also publish comments on original papers that have been published elsewhere?

invariably translated into an increased supply of replication studies. In economics, the *Journal of Political Economy* and *Labor Economics* discontinued formats dedicated to replication due to a lack of submissions (Hamermesh 2007). Likewise, the *American Economic Review* features an on-line format for replications with only one published comment, according to the editors due to a lack of submissions as well (see Ankel-Peters, 2023b). Hence, there are reasons to believe that the dearth of replications in journals is also a supply problem, probably because scholars have little incentives, intrinsic or extrinsic, to engage in reproductions and replications.

## Generating and promoting reproductions and replications

Several authors of this article founded the Institute for Replication to address the above issues by promoting and generating reproductions and replications on an ongoing basis. I4R's main goals are to assess and improve the computational reproducibility of research and its replicability.

As of 2023, I4R reproduces and replicates studies published in the *American Journal of Political Science*, the

*American Political Science Review*, and the *Journal of Politics*. I4R recruits replicators for studies published in 2022 onward in those outlets. I4R's team or a data editor computationally reproduces the numerical results, allowing replicators to focus on robustness reproductions and replicability.

To assist with the recruitment of replicators, I4R set up a board of editors from various research fields and with various institutional ties, thus allowing it to cast a very wide net. An editor's task is specifically to identify potential replicators. The institute currently has a board of editors for economics, finance, and political science, which actively recruits and selects replicators for studies recently published in top journals in each field. Of note, replicators may be faculty members or graduate students.

I4R also recently developed replication games to generate reproductions and replications in political science. Replication games are meet-ups ("hackathons") open to faculty, post-docs, graduate students, and other researchers. Participants join a small team and are asked to first computationally reproduce, then to carry out additional reproducibility analyses of a published paper or study in their field of interest. In practice, teams work during the event and the following weeks on testing the robustness of the results of a prior study using the same data but different analytical



decisions than made by the original investigator. All replication reports are then combined into (mega) meta-papers, and all replicators are offered co-authorship.

### *Reproducibility and replicability in class*

We argue that reproduction and replication of research by graduate students plays a pivotal role in upholding the integrity and credibility of scientific inquiry, laying the foundation for the advancement of knowledge. Reproducing and replicating the work of others is a fundamental and essential aspect of graduate education (Janz, 2016). Every year, students, and more generally, researchers around the world, carry out reproduction exercises, generating important pieces of new knowledge. Unfortunately, those reproductions and replications exercises are rarely publicly documented or rewarded. One recent platform developed to deal with this issue is the Social Science Reproduction Platform (SSRP). This resource standardizes and crowdsources exercises of computational reproducibility, and provides extensive guidance on how to carry out a reproduction exercise. First, students and researchers typically verify the existence of reproduction materials for an article. Second, they assess how reproducible these materials are. Third, they might make some improvements to these materials (from fixing file paths and libraries, to translating code into a different programming language). Finally, they often explore different specifications to see which results may or may not robustly hold.

### *How to make adversarial exchanges more collaborative*

Reproductions and replications in academia can sometimes become adversarial. The process can potentially lead to tensions between the replicators and the original authors if the replication study fails to replicate the original findings (Laitin and Reich, 2017). This can occur due to various reasons such as differences in sample characteristics, variations in experimental conditions, or even methodological limitations of the replication study itself. When the replication results contradict the original findings, it may challenge the credibility and impact of the original study, leading to a defensive response from the original authors.

Another factor that can contribute to adversarial relations is that the original authors may perceive the replication as an attempt to undermine their work, and as a result, may respond defensively or dismissively, seeking to protect their intellectual contributions.

To mitigate adversarial dynamics, fostering open communication, transparency, and collaboration between replicators and original authors is crucial. I4R, for instance,

deals with communication between original authors and replicators (Brodeur, 2023). By acting as an intermediary between authors and replicators, it helps researchers collectively contribute to a more robust and reliable body of knowledge and makes exchanges less adversarial. An additional approach for enhancing the efficiency of conflict resolution might involve embracing the framework of adversarial collaborations as proposed by Kahneman and Klein (2009) which are increasingly being used in different areas of the social sciences (e.g., Clark and Tetlock, 2023).

### *Incentives for replicators*

Engaging in replication studies can carry potential negative consequences for the career of replicators. These consequences can arise from a variety of factors, including the prevailing limited incentives.

First, replicators may find it challenging to gain recognition and visibility for their work, as replication studies may be difficult to publish and are often less recognized by their peers. Additionally, replication studies can elicit negative reactions from original authors, as discussed earlier. If replicators challenge or refute the original findings, they may face criticism, or even personal attacks from the authors or their supporters. These adversarial interactions can create a hostile environment for replicators and potentially damage their professional relationships within the academic community.

Second, dedicating time and resources to replication studies may divert replicators' attention from pursuing their original research agendas. The time spent replicating studies and addressing potential challenges can slow down their career progression and limit their ability to build a unique research portfolio.

To mitigate some of these negative consequences, we make the case for a growing recognition of the importance of replication studies within political science. One solution is to combine replications into large meta-papers. Being granted co-authorship to a meta-paper encourages researchers to replicate studies and changes incentives and the way replication is conducted. This is in part due to the fact that replicators work in teams and are not as pressured to show that the original findings are not replicable or robust. Moreover, meta-papers allow for estimating a replication rate within a discipline or subfield. Inferring the replication rate from published one-on-one reproductions and replications is not possible since problematic reproductions and replications are more inclined to be conducted in the first place and also more likely to be published. Last, one key editorial policy at I4R is that replicators may remain anonymous and still get co-authorship on a meta-paper.

However, the identity of the replicator is known to the editorial board that vetted this person and their work.

## Conclusion and recommendations

Leading political science journals have recently adopted innovative open science practices, incorporating policies that emphasize the availability of data and code, along with the inclusion of reproducibility analysts. We believe more journals ought to implement similar policies in a way that does not stifle creativity and that minimizes excess burden for researchers, editors and journal staff.

A key question going forward, which we have not addressed is “Which papers should be replicated?” We believe greater reflection is warranted on this matter. Interesting options include crowd forecasts to determine which papers are likely to run into replication failures. Crowdsourcing may also be helpful in prioritizing the papers that most demand replication (either because the papers are very important or because their findings are especially dubious).

We urge researchers, journal editors, and funders to start holding political science to higher open science standards, and supporting and facilitating the conduct and publication of replication and reproducibility studies. We make three recommendations:

- (1) We call on the *American Political Science Association* sections (or at least the experiments section) to create an award for best replication/extension study.
- (2) We urge the creation of an outlet dedicated to replications backed up by one of the large disciplinary professional associations (for which impact factor might not be a primary consideration).
- (3) We also recommend that more political science journals start using data editors to improve computational reproducibility.

## Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Abel Brodeur  <https://orcid.org/0000-0003-3980-4324>

Kevin Esterling  <https://orcid.org/0000-0002-5529-6422>

Matthew Hepplewhite  <https://orcid.org/0000-0003-0436-578X>

Magnus Johannesson  <https://orcid.org/0000-0001-8759-6393>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Ankel-Peters J, Fiala N and Neubauer F (2023a) *Is economics self-correcting? Replications in the American economic review*. Ruhr Economic Papers No. 1005.
- Ankel-Peters J, Fiala N and Neubauer F (2023b) Do economists replicate? *Journal of Economic Behavior & Organization* 212: 219–232.
- Arel-Bundock V (2022) *Quantitative Political Science Research Is Greatly Underpowered IAR Discussion Paper Series*. No. 6.
- Avelino G, Desposato S and Mardegan I (2021) Transparency and Replication in Brazilian Political Science: A First Look. *Dados* 64(3).
- Blair G (2019) Declaring and diagnosing research designs. *American Political Science Review* 113(3): 838–859.
- Breznau N (2022) Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences* 119(44): e2203150119.
- Brodeur A (2016) Star wars: the empirics strike back. *American Economic Journal: Applied Economics* 8(1): 1–32.
- Brodeur A, Cook N and Heyes A (2020) Methods matter: P-hacking and publication bias in causal analysis in economics. *The American Economic Review* 110(11): 3634–3660.
- Brodeur A (2023) Replication games: how to make reproducibility research more systematic. *Nature* 621(7980): 684–686.
- Bryan CJ, Yeager DS and O’Brien JM (2019) Replicator degrees of freedom allow publication of misleading failures to replicate. *Proceedings of the National Academy of Sciences* 116(51): 25535–25545.
- Bueno de Mesquita B (2003) Symposium on replication in international studies research. *International Studies Perspectives* 4(1): 72–107.
- Camerer CF (2016) Evaluating replicability of laboratory experiments in economics. *Science* 351(6280): 1433–1436.
- Camerer CF (2018) Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour* 2(9): 637–644.
- Chang AC and Li P (2022) Is economics research replicable? Sixty published papers from thirteen journals say “often not”. *Critical Finance Review* 11(1): 185–206.
- Christensen G and Miguel E (2018) Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature* 56: 920–980.
- Clark CJ and Tetlock PE (2023) Adversarial collaboration: the next science reform. In: CL Frisby, RE Redding, WT O’Donohue, et al. (eds) *Ideological and Political Bias in Psychology*:

- Nature, Scope, and Solutions*. Cham: Springer International Publishing, 905–927.
- Clemens MA (2017) The meaning of failed replications: a review and proposal. *Journal of Economic Surveys* 31: 326–342.
- Coffman LC, Niederle M and Wilson AJ (2017) A proposal to organize and promote replications. *American Economic Review* 107(5): 41–45.
- Colliard JE (2021) *The Economics of Research Reproducibility*. HEC Paris Research Paper No. FIN-2019-1345.
- Dafoe A (2014) Science deserves better: the imperative to share complete replication files. *PS: Political Science & Politics* 47(1): 60–66.
- Dreber A and Johannesson M (2023) *A Framework for Evaluating Reproducibility and Replicability in Economics*. I4R Discussion Paper Series No. 38.
- Esterling KM (2023) *The necessity of construct and external validity for generalized causal claims*. I4R Discussion Paper Series. No. 18.
- Findley MG (2021) External validity. *Annual Review of Political Science* 24: 365–393.
- Gerber A and Malhotra N (2008) Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Quarterly Journal of Political Science* 3(3): 313–326.
- Gertler Galiani PS and Romero M (2018) How to make replication the norm. *Nature* 554(7693): 417–419.
- Gleditsch NP and Janz N (2016) Replication in international relations. *International Studies Perspectives* 17(4): 361–366.
- Hamermesh DS (2007) Viewpoint: replication in economics. *Canadian Journal of Economics* 40: 715–733.
- Ioannidis JP, Stanley TD and Doucouliagos H (2017) The power of bias in economics research. *Economic Journal* 127(605): 236–265.
- Janz N (2016) Bringing the gold standard into the classroom: replication in university teaching. *International Studies Perspectives* 17(4): 392–407.
- Kahneman D and Klein G (2009) Conditions for intuitive expertise: a failure to disagree. *American Psychologist* 64(6): 515.
- Kerr NL (1998) HARKing: Hypothesizing after the results are known. *Personality and social psychology review* 2(3): 196–217.
- King G (1995) Replication, replication. *PS: Political Science & Politics* 28(3): 444–452.
- Laitin DD and Reich R (2017) Trust, transparency, and replication in political science. *PS: Political Science & Politics* 50(1): 172–175.
- Lin W and Green DP (2016) Standard operating procedures: a safety net for pre-analysis plans. *PS: Political Science & Politics* 49(3): 495–500.
- McCullough BD, McGeary KA and Harrison TD (2006) Lessons from the JMCB archive. *Journal of Money, Credit, and Banking* 38(4): 1093–1107.
- Mueller-Langer F, Fecher B, Harhoff D and Wagner GG (2019) Replication studies in economics—How many and which papers are chosen for replication, and why? *Research Policy* 48(1): 62–83.
- Open Science Collaboration. (2015) Estimating the reproducibility of psychological science. *Science* 349(6251): aac4716.
- Peters J, Langbein J and Roberts G (2018) Generalization in the tropics—development policy, randomized controlled trials, and external validity. *The World Bank Research Observer* 33(1): 34–64.
- Simmons JP, Nelson LD and Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11): 1359–1366.
- Young C and Holsteen K (2017) Model uncertainty and robustness: a computational framework for multimodel analysis. *Sociological Methods & Research* 46(1): 3–40.