

A Proposal to Organize and Promote Replications[†]

By LUCAS C. COFFMAN, MURIEL NIEDERLE, AND ALISTAIR J. WILSON*

A number of recent results in peer-reviewed academic work have ignited a debate over the fraction of results that are false positives. Published work in top economics journals has discontinuities around p -values of 0.05, consistent with researcher data mining (Brodeur et al. 2016). Moreover, in an attempt to replicate a set of recent experimental articles, Camerer et al. (2016) were able to successfully replicate 61 percent of the studies, substantially less than the original p -values would suggest.

One response to such findings is to minimize flexibility over the econometric specifications available to researchers via registration and pre-analysis plans.¹ However, the sheer volume of research, coupled with the strong tendency only to finish and publish positive findings (Franco, Malhotra, and Simonovits 2014) leads to false positives even in the absence of questionable research practices (Ioannidis 2005; Coffman and Niederle 2015; Simonsohn, Nelson, and Simmons 2014). A solution is required to try and separate what is true from what is not after a paper is disseminated: replication.

Incentives for running replications are low. Replication projects lack novelty and can fail to attract interest. Further, attempting to replicate someone's work might make enemies, no matter the results. The upshot of all this is a lack of replications in economics. As we discuss in the next section, only the most famous papers are replicated with any regularity.

The other less obvious problem is sometimes even when replications do exist, they often take

forms that make them hard to find. Consider, for example, Niederle and Vesterlund (2007). Of the first two dozen independent replications (that is, without a coauthor of the initial study) only one can be found searching variants of “replication” along with “Niederle and Vesterlund (2007).” This is because most of these replications are embedded within a paper with a wider scope; the replication attempt is often a side result, or simply framed as a baseline treatment.² This is not uncommon. In the next section, we highlight recent work showing a majority of the current replication attempts in economics are hidden, and that this leads to authors and experts not knowing what work has been replicated and what has not.

In light of these two unfortunate facts about replications, we make a two-pronged proposal for strengthening the incentives for new replications as well as better organizing the replications that are conducted. First we propose that top journals include one-page “replication reports.” One type of report would be new work that shows whether a specific published paper was replicated (or not). A second type would consist of authors re-publishing an existing replication attempt that was already published elsewhere as part of a larger paper (such as the early replications of Niederle and Vesterlund). Second, we propose a norm of citing replication work alongside the original. Whereas the current norm for citing might simply be “see Roth and Kessler (2010),” our proposal would simply be to add, for example, “replicated by Camerer et al. (2016).”

The aim of the proposal is to increase both the visibility of current replications and the incentives to carry out new replications, since the costs can be substantial (see Hamermesh 2017 for additional discussions). Coffman and Niederle (2015) suggest that even a small number of replications can be very valuable to

*Coffman: Harvard University, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138 (e-mail: coffman@fas.harvard.edu); Niederle: Stanford University, 579 Serra Mall, Stanford, CA 94305, NBER and SIEPR (e-mail: niederle@stanford.edu); Wilson: University of Pittsburgh, 4900 Posvar Hall, 230 S. Bouquet Street, Pittsburgh, PA 15224 (e-mail: alistair@pitt.edu).

[†]Go to <https://doi.org/10.1257/aer.p20171122> to visit the article page for additional materials and author disclosure statement(s).

¹See Olken (2015) and Coffman and Niederle (2015) for discussions of pre-analysis plans.

²See Niederle (2016) for details.

increase the posterior that a specific publication presents a robust finding, and so slight shifts in the incentives and organization have the possibility of producing substantial effects.

I. The State of Replications

There are many ways in which we can replicate a result or provide some idea of its robustness. While what constitutes a replication or a robustness check may differ by field, we will broadly summarize these efforts when we use the term “replication” herein—including both positive and negative replications.

Large-scale replication exercises exist and are published in prominent outlets; this just happens at low frequencies. Camerer et al. (2016) published their replication of 18 prominent experimental papers in *Science*. The *Journal of Money Banking and Credit's* macroeconomic replication exercise was documented in the *American Economic Review* (Dewald, Thursby, and Anderson 1986). Moreover, the publication of papers from both this session and a sister session (Duvendack, Palmer-Jones, and Reed 2017; Höeffler 2017; Anderson and Kichkha 2017; Chang and Li 2017) illustrate a desire to address the topic.

However, outside of these infrequent efforts, economics journals are in general reluctant to publish explicit replication attempts. Particularly in top journals, our profession places much greater emphasis on novelty than documenting robustness. Focusing only on studies where the main purpose was validating a previous empirical result, Duvendack, Palmer-Jones, and Reed (2017) find that in the past 50 years the *American Economic Review* (*AER*) has published just 28 studies where the main purpose was validating an empirical result from a previously published study. Noticeably, in an earlier work (Duvendack, Palmer-Jones, and Reed 2015) they report that out of 15 of these *AER* replications, every single one was a negative result that failed to confirm the original. The publication incentives for pure replication exercise are skewed against positive replications, diminishing the incentives to carry them out, and, because of selection bias, the ability to update beliefs when they are.

Outside of explicit replications, the profession is doing slightly better, though there is substantial room for improvement. Examining

the *AER's* centenary volume, Berry et al. (2017) find that five years after publication, only 29 percent of empirical papers have a published replication attempt. This result is despite using a broad definition of replication, so that even papers with proprietary data could have been replicated (17 of the 70 papers in their sample did not make data available). Moreover, they find substantial variation in the incidence of replication: restricting attention to behavioral/experimental papers that rate is almost 50 percent, where it is 20 percent for the remaining fields. For papers with citation counts above the median for their field, 42 percent have replications, compared to 13 percent for those below the median.³ Using a similarly broad definition, Hamermesh (2017) shows that for ten of the most-cited papers in labor economics from the early 1990s, eight have been replicated at least four times, though two papers only have one and two replications, respectively.

Finally, even when replications occur, it does not enter into our collective consciousness. Both authors and experts in the field are often very unsure about whether a specific paper has been replicated (Berry et al. 2017). As mentioned, one possible reason is that many replications are hidden within bigger papers. Of the 52 papers that replicate a result in the centennial issue of the *American Economic Review*, all but eight are contained in wider-scope papers, where the replication was not the main focus (Berry et al. 2017). Further the large majority of coded replications (34 papers of 52) do not explicitly mention whether their results corroborate those of the cited volume paper, and so more cursory searches of these literatures would not detect the replications.

II. A Proposal to Organize and Increase the Visibility of Replications

We would like to promote new replications through greater incentives to researchers, and to better organize extant replications and make their results more visible. Ideally, a highly visible journal, such as the *AER*, would introduce a short, new section. In addition to *Shorter*

³Sukhtankar (2017) finds about 5 percent of recent top development papers have been replicated, though the rate is significantly larger for well-cited papers.

Papers, Comments, and Replies, there would be the *Replications* section. Publications within this section would be short: identifying the replicated paper, briefly describing any qualitative differences in procedures (if any), and the main conclusion. While the published replication paper would be an extended abstract, a more extensive online Appendix with data, code, and formal tests would be linked for interested readers.

There are two reasons why this would increase the visibility and the availability of replications: First, it would allow authors who wrote and published a paper that builds on and extends an important result to re-publish aspects of their data that specifically replicates it. While there are some costs in writing the replication paper, the incentive of an additional publication might be sufficient to encourage the endeavor. Such an exercise would make trivial the task of finding “hidden” replications. Where the extensive coding in Berry et al. (2017), Hamermesh (2017), and Sunkhtankar (2017) took many hours of work to discover and code replications, our proposal would allow for discovery through a simple indexed search. Moreover, this process would de-bias the search for replications. Positive and negative replications would receive equal weight, where casual empiricism suggests that presently negative replications are more visible as they have a higher chance of publication.

Such a *Replications* section could also encourage and illuminate less formal replication attempts that may not be written up otherwise. For example, in many graduate programs, students still learning technical skills are often asked to replicate famous papers’ results. While these replications are occasionally published, this is more often the case when they fail to replicate the original paper, and even then publication often presents a big challenge to young researchers. It would be equally valuable to know which papers are replicated, to get a better sense on how the number of false positives.

Once a desired number of replications were written, one can imagine that with many positive replications a coda meta-analysis could be written—including all the authors of replications—to summarize the findings. Likewise, in the case of many failed replications. In the case of mixed results, more replications would be encouraged. That is with replications being visible, it would

be clearer which papers still needed additional replication, while others may be deemed to be robust given the large evidence of positive replications (for example, after a meta-analysis was written). Replication efforts would therefore be directed toward papers that had not been replicated, as this would increase the chances of the work being published.

The second part of the proposal provides an incentive to produce replication work through the currency of our industry: citations. While it may be too costly for top journals to publish all replications, journals could encourage citation systems where citations to the original paper include citations to its replications. This would only cost a few extra lines, and would ensure that well-executed replications receive credit. A second benefit would be to the journal. If replications of especially highly cited papers are welcomed and published, and those replications are cited alongside the original article, this would ensure that the replication reports would have above-average citation counts, and hence not reduce the impact factor of the journal.⁴

One additional important question is to assess which papers are of sufficient importance for a replication to be desirable. While clearly there should be editorial freedom in making that decision, agreement on a list of important papers to be replicated would help focus replication efforts. The selection of such papers would ideally focus on well-cited papers. One possibility might be to focus on the top half most-cited papers (by field) published in top five journals alongside very highly cited papers published in other outlets.

These changes might necessitate the introduction of a new JEL code for replications. This, along with the possible benefit of citations, would most likely help papers self-identify the components of their works that are replications. This would thereby reduce the burden of writing overview articles, and finding replication attempts.

This proposal neglects many details. What constitutes a good and fair replication? How do we avoid unfair targeting of the projects that get replicated? Who should be making editorial

⁴ Admittedly, both impact factors and the productivity of researchers would have to take into account whether citations are for original or replication works.

and reviewing decisions? Given the space constraints of this article, we simply put forward a noncontroversial first step, one that we think many can agree on. But we recognize the difficulty in answering these questions (among many others) more fully.

Finally, we believe these changes need to come from the top down.⁵ Replication work, even if published, needs visibility to affect awareness; it needs the visibility afforded by top journals. Further, the norms of citing replication work will only be solidified if it comes in well-read papers. This can only happen if done by well-published authors and enforced by editors at top journals.⁶

Many researchers are concerned by the recent upheaval in psychology determining that most papers cannot be replicated. Clearly the issue of replicability is a source for substantial concern in economics as well.⁷ While we believe the economics profession is not doing too badly on the dimension of replicability, this is an area that economists can be leaders in designing better mechanisms for promulgating academic research.

REFERENCES

- Anderson, Richard, and Areerat Kichkha.** 2017. "Replication, Meta-Analysis, and Research Synthesis in Economics." *American Economic Review*. <https://doi.org/10.1257/aer.p20171033>.
- Berry, James, Lucas C. Coffman, Rania Gihleb, Douglas Hanley, and Alistair J. Wilson.** 2017. "Assessing the Rate of Replication in Economics." *American Economic Review*. <https://doi.org/10.1257/aer.p20171119>.
- Brodeur, Abel, Mathias Le, Marc Sangnier, and Yanos Zylberberg.** 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* 8 (1): 1–32.
- Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al.** 2016. "Evaluating Replicability of Laboratory Experiments in Economics." *Science*.
- Chang, Andrew C., and Andrew Li.** 2017. "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not'." *American Economic Review*. <https://doi.org/10.1257/aer.p20171034>.
- Coffman, Lucas C., and Muriel Niederle.** 2015. "Pre-analysis Plans Have Limited Upside, Especially Where Replications Are Feasible." *Journal of Economic Perspectives* 29 (3): 81–98.
- Dewald, William G., Jerry G. Thursby, and Richard G. Anderson.** 1986. "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project." *American Economic Review* 76 (4): 587–603.
- Duwendack, Maren, Richard W. Palmer-Jones, and W. Robert Reed.** 2015. "Replications in Economics: A Progress Report." *Econ Journal Watch* 12 (2): 164–91.
- Duwendack, Maren, Richard Palmer-Jones, and Robert Reed.** 2017. "What is Meant by 'Replication' and Why Does It Encounter Such Resistance in Economics?" *American Economic Review*. <https://doi.org/10.1257/aer.p20171031>.
- Franco, Annie, Neil Malhotra, and Gabor Simonovits.** 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502–05.
- Hamermesh, Daniel S.** 2017. "Replication in Labor Economics: Evidence from Data and What It Suggests." *American Economic Review*. <https://doi.org/10.1257/aer.p20171121>.
- Höeffler, Jan H.** 2017. "Replication and Economics Journal Policies." *American Economic Review*. <https://doi.org/10.1257/aer.p20171032>.
- Ioannidis, John P. A.** 2005. "Why Most Published Research Findings Are False." *PLoS Med* 2 (8): e124.
- McCullough, B. D.** 2009. "Open Access Economics Journals and the Market for Reproducible Economic Research." *Economic Analysis and Policy* 39 (1): 117–26.
- Niederle, Muriel.** 2016. "Gender." In *The Handbook*

⁵One prestigious first-mover can lead to shifting norms across the profession. Consider the *AER*'s adoption of open-data policies in 2003, after which they were adopted by *Econometrica*, the *Review of Economic Studies*, and the *Journal of Political Economy* (see McCullough 2009 for a timeline).

⁶As discussed in Nosek, Spies, and Motyl (2012), new replication or negative-result journals are a nonstarter. Without the prestige of a large professional association or a high-impact journal to begin with, such journals are doomed to failure.

⁷The Replication Network (<http://replicationnetwork.com>) whose aim is to promote replications and to ultimately have economics journals publish those replications, had been endorsed by 350 economists at the time of writing.

- of Experimental Economics*, Vol. 2, edited by John H. Kagel and Alvin E. Roth, 481–562. Princeton: Princeton University Press.
- Niederle, Muriel, and Lise Vesterlund.** 2007. “Do Women Shy Away from Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics* 122 (3): 1067–1101.
- Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl.** 2012. “Scientific Utopia II. Restructuring Incentives and Practices to Promote Truth over Publishability.” *Perspectives on Psychological Science* 7 (6): 615–31.
- Olken, Benjamin A.** 2015. “Promises and Perils of Pre-analysis Plans.” *Journal of Economic Perspectives* 29 (3): 61–80.
- Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons.** 2014. “P-curve: A Key to the File-Drawer.” *Journal of Experimental Psychology: General* 143 (2): 534–47.
- Sukhtankar, Sandip.** 2017. “Replications in Development.” *American Economic Review*. <https://doi.org/10.1257/aer.p20171120>.