INSTITUTE for REPLICATION

No. 13
I4R DISCUSSION PAPER SERIES

Do Economists Replicate?

Jörg Ankel-Peters

Nathan Fiala

Florian Neubauer

January 2023

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 13

Do Economists Replicate?

Jörg Ankel-Peters[1], Nathan Fiala[2], Florian Neubauer[2]

[1]RWI – Leibniz Institute for Economic Research and University of Passau
[2]University of Connecticut and RWI – Leibniz Institute for Economic Research

JANUARY 2023

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Do Economists Replicate?

JÖRG ANKEL-PETERS, NATHAN FIALA & FLORIAN NEUBAUER*

January 06, 2023

Reanalyses of empirical studies and replications in new contexts are important for scientific progress. Journals in economics increasingly require authors to provide data and code alongside published papers, but how much does the economics profession actually replicate? This paper summarizes existing replication definitions and reviews how much economists replicate other scholars' work. We argue that in order to counter incentive problems potentially leading to a replication crisis, replications in the spirit of Merton's 'organized skepticism' are needed – what we call 'policing replications'. We review leading economics journals to show that policing replications are rare and conclude that more incentives to replicate are needed to reap the fruits of rising transparency standards.

Keywords: replication, replicability, research transparency, meta-science, generalizability, systematic review

JEL-Classifications: A11, C18

*Ankel-Peters: RWI and University of Passau (e-mail: peters@rwi-essen.de). Fiala: University of Connecticut and RWI (e-mail: nathan.fiala@uconn.edu); Neubauer: University of Connecticut and RWI (e-mail: florian.neubauer@uconn.edu); Funding: This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) [Grant No. 3473/1-1 within the DFG Priority Program META-REP (SPP 2317)]. We thank Gunther Bensch, Abel Brodeur, Daniel Hamermesh, David McKenzie, Frank Müller-Langer, Tim Salmon, Maximiliane Sievert, Sandip Sukhtankar, Séverine Toussaert, Alistair Wilson, seminar participants at ZBW, and two anonymous reviewers for valuable comments and suggestions. We are particularly grateful to Niklas Benner of RWI for supporting us with textual analysis techniques to identify citations in the abstracts of papers published in the Top 50 journals. Caitlin Chan provided valuable research assistance.

1. Introduction

Replications are important in all empirical disciplines to verify results, uncover errors and negligence, and test the generalizability of previous findings to new contexts. This is especially true for empirical economics with its profound implications for policy decisions. Other disciplines such as psychology and medical science have gone through intense debates about replication crises, based on cases of fraud and large-scale replication projects (Bryan et al., 2019; Hensel, 2021; Maxwell et al., 2015; O'Grady, 2020; Open Science Collaboration, 2015; Piller, 2022; Servick, 2018; Simonsohn, 2013). Within economics, a growing recent literature raises concerns about the replicability of empirical findings (Brodeur et al., 2020, 2016; Camerer et al., 2016; Christensen and Miguel, 2018; Ferraro and Shukla, 2022, 2020; Huntington-Klein et al., 2021; Ioannidis et al., 2017; Peters et al., 2018; Vivalt, 2020), suggesting that the economics profession might need some introspection in the spirit of Robert Merton's norm of 'organized skepticism' (Merton, 1973). An essential component of organized skepticism, on top of peer review, we argue, are replications that challenge published studies. In the present paper, we examine how widespread replication is in economics.[1]

First, we review the three existing systematic reviews that estimate how often replications are being conducted in economics: Berry et al. (2017), Mueller-Langer et al. (2019), and Sukhtankar (2017). They provide widely diverging replication rates and, as we will show, the most important reason for this is different definitions of what constitutes a replication. We contend that all definitions used in the three reviews are reasonable but differ in how directly they address previously published results. Broader definitions of replications include studies that build on an existing empirical finding by slightly modifying the research question and applying it in a new context.

[1] To avoid terminological confusion: throughout the paper we use the terms 'replication' and 'replicate' to refer to the *process* of replication (e.g. through a reanalysis or robustness replication), not to a succesful attempt of obtaining the same result as the previously published study.

Such *implicit* replication work is daily fare in economics and, hence, applying such broader definitions delivers high replication rates.

While we acknowledge the scientific importance of implicit replications, we argue that for organizing skepticism in Merton's spirit a narrower type of replication is needed. Such replications should stress-test published results to uncover purposeful or unintentional questionable research practices, as they have been diagnosed, for example, in Ferraro & Shukla (2020, 2022). We refer to this as *policing replication*. In most cases, questionable research practices are not "*blatantly improper*" but "*offer considerable latitude for rationalization and self-deception*" (John et al., 2012). They comprise, among others, *p*-hacking (Brodeur et al., 2022a, 2022b, 2022c, 2020; Ferraro and Shukla, 2020; Huntington-Klein et al., 2021; Kranz and Pütz, 2022), ex-post theorizing (Kerr, 1998), reporting underpowered results (Dahal and Fiala, 2020; Ioannidis et al., 2017), uncorrected multiple hypothesis testing (Anderson, 2008; Fink et al., 2014), and coding errors (Foote and Goetz, 2008). The 'policing replication' category complements existing replication definitions in other works such as Clemens (2017), Dreber and Johannesson (2022), and Hamermesh (2007), see Table 1 for an overview. The constitutive feature of a replication to qualify as policing, we argue, is the direct engagement with the original work in the title or the abstract. In this sense, all sub-categories of existing replication taxonomies, a priori, qualify as policing.

In a next step, we examine how frequent policing replications are in economics. We use three approaches that deliver a coherent picture. First, we review the Top 50 economics journals for policing replications and find that 259 (or 0.9%) of all 29,682 published papers in the Top 50 economics journals between 2010 and 2020 fit our definition. Second, we corroborate this finding by isolating those replications from Berry et al. (2017), Sukhtankar (2017), and Mueller-Langer et al. (2019) that would qualify as policing replications. Third, we look at how many comments have been published in the *American Economic Review (AER)* over time, one of the profession's leading journals. Comments in the *AER* discuss and challenge papers that were previously published in the journal. To the extent that these comments are empirical

(some are theoretical), they are in line with our understanding of policing replications. We find that there has been a continuous downward sloping trend over the past decades that is even more noteworthy given the sharp increase in empirical papers in the *AER*. In recent years, less than 3 % of papers published in the *AER* were comments. Finally, we conducted a short survey among the editors of the Top 50 economics journals asking them whether they publish replications that meet the policing requirement: 79% of the responding editors stated that they publish replications of papers originally published in their journal, and of those, 62% indicated that they also publish replications of papers published in other journals.

Our paper contributes to a growing meta-scientific literature in economics and its research transparency debate (Christensen and Miguel, 2018). Important reviews demonstrate that the economics profession has made tremendous progress in recent years on the availability of data and code for published work (Christensen et al., 2020; Christensen and Miguel, 2018; Miguel, 2021; Vilhuber, 2020a). While stricter data sharing policies as such already encourage better scientific conduct (see Askarov et al., 2022), we argue that deep and sustainable incentives for credible research only emerge if, in parallel, a replication culture is established that takes the data to the test (Höffler, 2017). Ofosu and Posner (2021) and Laitin (2013) make a similar case for the effectiveness of pre-analysis plans (PAP) in combating *p*-hacking and publication bias. As Ofosu and Posner (2021) point out: "*whatever the benefits of pre-registration may be in theory, PAPs are unlikely to enhance research credibility without vigorous policing*".

2. Replication definitions and replication rates in economics

There is no universally accepted definition of replication in economics. Table 1 summarizes the prevailing definitions and sub-categories along three dimensions that demarcate the different replication types: whether the new paper uses the same specification, the same population, and the same sample as the original paper. The most influential papers that define replication types in economics are Hamermesh (2007) and Clemens (2017), as well as an overview paper on research transparency by

Christensen and Miguel (2018). Another compelling categorization is Freese and Peterson (2017) from quantitative sociology, a sister discipline. More recently, two initiatives have gained momentum in the social sciences community, the *Berkeley Initiative for Transparency in the Social Sciences* (*BITSS*) and the *Institute for Replication* (*I4R*), which both host platforms to crowdsource computational and robustness replications. We include the nomenclature used by the *I4R* in Table 1; BITSS' categorization combines the Hamermesh (2007) and the Clemens (2017) definitions.

Table 1: Most significant replication definitions in the social sciences

Author(s), Year	Category	New paper uses the same…		
		Specification[1]	Population	Sample
Institute for Replication; Dreber and Johannesson (2022)	Computational Reproduction	✓	✓	✓
	Recreate Reproduction[2]	✓/✗[3]	✓	✓
	Robustness Replication	✗	✓	✓
	Direct Replication	✓	✓/✗[4]	✗
	Conceptual Replication	✗	✓/✗[4]	✗
Clemens (2017)	Verification	✓	✓	✓
	Reproduction	✓	✓	✗
	Reanalysis	✗	✓	✓/✗[5]
	Extension	✓	✗	✗
Freese and Peterson (2017)	Verifiability	✓	✓	✓
	Robustness	✗	✓	✓
	Repeatability	✓	?[6]	✗
	Generalization	✗	?[6]	✗
Hamermesh (2007)	Pure replication	✓	✓	✓
	Statistical replication	✓	✓	✗
	Scientific replication	✓/?[7]	✗	✗

Notes: [1]Hamermesh (2007) used the term 'model', and Freese and Peterson (2017) vary in their wording between 'analysis', 'specification', 'procedure', and 'method'. [2]Dreber and Johannesson (2022) introduce this additional category which differs from "computational reproduction" only in that it emphasizes the usage of raw data and not having the analysis code of the original paper. This category is not included in the *I4R* definition. [3]The specification in the reproduction is not always identical to the original paper as the replicator does not have access to the original code but tries to recreate the analysis based on the given information in the original paper. [4]*I4R*'s definitions of direct and conceptual replication only require new data but it does not matter if it is from the same population or not. Dreber and Johannesson (2022) further subdivide between the same, similar, and different populations. [5]According to Clemens, a reanalysis can use exactly the same data as the original study or a new sample from the same population. [6]Freese and Peterson (2017) do not specify whether the data shall come from the same or a new population. [7]Similar but not identical specification.

We next turn to replication rates in economics. Three recent reviews have systematically addressed how much economists replicate, summarized in Table 2.[2]

[2] Hensel (2021) compares replication rates in various management-related disciplines to those in psychology and economics, referring to the rates for economics found in Mueller-Langer et al. (2019) and Berry et al. (2017) as the lower and upper bound.

Table 2: Overview of papers investigating replication rates

Paper	Replication rates	Definition of replication	Search engine	Inclusion criteria	Search strategy for replications, coding
A. Selective Replications - "How many published papers are replicated?"					
Sukhtankar 2017	Overall (incl. working papers): 6.2% Published: 3.3% RCTs: 12.5% 71 replication studies were found, and they include: Replication verification: 32.4% Replication extension: 0% Robustness reanalysis: 77.5% Robustness extension: 36.6% (they don't add up to 100% because some studies included different replication types)	Clemens nomenclature[1] I. Replication: a) Verification b) Reproduction II. Robustness: c) Reanalysis d) Extension	GS[3]	Original papers: - Top five journals (AER, Econometrica, JPE, QJE, ReStud) and next five general-interest journals: AEJ:AE, AEJ:EP, EJ, JEEA, ReStat - JEL code: O - 2000-2015 Replicating papers: Published, working papers	- First step: Formalized word search in GS search among citing papers for "replicate OR replicates OR replicated OR replication OR replicating" - Second step: Subjective coding of replications, i.e., no formalized criteria or protocol in decision whether a paper is a replication or not - Supplemented GS search with search on other websites (see column "additional sources")
Berry et al. 2017	Replication: 28.6%[2] Extension: 48.6% Robustness: 40% Any of the three: 60%	Definitions of rates in previous column: A. 'Replication': "Any project that reports results that speak directly to the veracity of the original paper's main hypothesis" B. 'Extension': "Testing a closely related hypothesis to the original paper" C. 'Robustness': Clemens' Robustness categories: Robustness reanalysis, Robustness extension	WoS	Original papers: - AER centenary volume (2010) Replicating papers: - Top 200 economics journals, Published papers only - 2010-2016	- Checked every citing paper of the 70 papers in the AER centenary volume whether it is a replication or not - Subjective coding of replications, i.e., no formalized criteria or protocol in decision whether a paper is a replication or not
B. Total Replications - "How many published papers are replications?"					
Mueller-Langer et al. 2019	0.1%	A. Narrow: Same data and code B. Wide replication: a) new data, same methods, same models b) same data, new methods, new models c) new data, new methods, new models	WoS	Original and replicating papers: - Top 50 Econ journals - Published papers - 1974-2014	- First step: Formalized word search in title and abstract for keywords such as “repli*,” “reexamin*,” “comment,” “revisit,” “retesting,” or “reappraisal” (among others), as well as references to other articles - Second step: Used frequency and location of keywords to determine likelihood of being a replication, then ranked them for each journal and looked at the 100 highest ranked papers in each journal in detail - Also included all eligible replications from ReplicationWiki in their dataset

Notes: Mueller-Langer et al. (2019) and Sukhtankar (2017) both used the replication database of the University of Göttingen as an additional source to find replications; Sukhtankar (2017) further used replicationnetwork.com and the 3ie Replication Paper Series. [1]See also Table 1. [2]These categories do not reflect the Clemens nomenclature, except for the "Robustness" category, which comprises Clemens' reanalysis and extension categories. See Section 2 for the definition of the categories used in Berry et al. (2017). [3]GS is Google Scholar. WoS is Web of Science.

The diagnosed replication rates range between 0.1% and 60%, which is the upper bound of Berry et al. (2017). The broad range can be ascribed to different definitions of 'replication rate'. To start with, the replication rate can measure two things: first, how many published papers *are replicated*. Or, second, how many published papers *conduct a replication*. We refer to the first as the *selective replication* approach, pursued by Berry et al. (2017) and Sukhtankar (2017) and also similar to the logic of Hamermesh (2017), who argues that only influential studies need to be replicated. We refer to the second approach as the *total replication* approach, used by Mueller-Langer et al. (2019), who review all papers in the Top 50 journals published between 1974 and 2014 to check how many of these are replications. The logic here is that replications should be part of the regular scientific process and hence also appear in journals.

The three reviews also use different definitions for what constitutes a replication in the first place. Berry et al. (2017) use a very pragmatic approach and define three categories: 'replication', 'extension', and 'robustness tests'.[3] A 'replication', according to Berry et al. (2017), is "*any project that reports results that speak directly to the veracity of the original paper's main hypothesis*" (p. 27). An 'extension' is a paper that is "*testing a closely related hypothesis to the original paper*" (p. 28). 'Robustness tests' are papers that either use the same specification in a new sample and population or different specifications on the same data. Berry et al. (2017) find that 28.6%, 48.6%, and 40% of papers in the *AER* volume under scrutiny are 'replications', 'extensions', and 'robustness tests', respectively. The authors emphasize in their abstract that 60% of papers in this *AER* volume have either a 'replication', 'robustness test', or an 'extension'. All three categories are very inclusive and broad.[4] Berry et al. (2017) also document narrower categories like 'verifications' and 'reproductions' using the Clemens (2017) definitions (see Table 1), for which they find zero and two cases, respectively.

[3] Note that Berry et al. (2017) deviate from the Clemens definition of these terms.

[4] For example, Berry et al. (2017) coded Magnan et al. (2015) as a 'replication' of the paper by Conley and Udry (2010), which looks at social learning in driving the adoption of fertilizer for pineapple production in Ghana. Five years later, Magnan et al. (2015) investigate how social learning affects the demand for a water-saving agricultural technology in India.

Sukhtankar (2017), the other review that uses the selective replication approach, applies a narrower definition, strictly following the Clemens (2017) nomenclature. Correspondingly, the overall replication rate in Sukhtankar (2017) is much lower at 3.3% when only replications in peer-reviewed journals are included and 6.2% when extended to replications published as working papers. Likewise, Mueller-Langer et al. (2019) apply a narrower definition, guided by the Hamermesh (2007) categories, but elicit the total replication rate of how many papers published in the Top 50 journals are replications. They find 130 replications in all 126,505 published papers between 1974 and 2014 (0.1%). The considerably higher replication rates in Berry et al. (2017) and Sukhtankar (2017) are probably in line with expectation due to their selective replication approach. Arguably, scarce replication work concentrates on more influential studies in selected top journals. Therefore, focusing on how often these papers are replicated, irrespective of where the replications are published, delivers higher rates than how often replications are published in the Top 50 journals.

3. Policing replication

3.1 A plea for more clarity: Assuming the burden-of-proof

We fully acknowledge the scientific value of *implicit* replications in the spirit of broader definitions as they are used, for example, in Berry et al. (2017). Yet, to organize the skepticism that Merton (1973) called for, *explicit* replications that directly scrutinize whether a paper's claim is valid are needed. We propose the term '*policing replications*' for this type of replication – building on Ofosu and Posner (2021), who coin the term '*policing'* in the context of checking PAPs against the actual analysis done in the empirical work they are supposed to pre-specify. The term was also used by Merton (albeit not in the context of his 'organized skepticism' norm): "*[…] the activities of scientists are subject to rigorous policing [by fellow experts], to a degree perhaps unparalleled in any other field of activity.*" (Merton, 1973, p. 276).

The deficiency we lament here is that for most replications in broader senses, it is effectively left to the reader to perceive a study as a replication or not (or to the coder,

as in the case of the three summarized reviews). We believe there should be more clarity about whether a new paper "*speaks to the veracity*" (Berry et al., 2017) of a previous study and we would therefore re-emphasize another important proposition of Clemens (2017): "*the burden of proof [for] a study to demonstrate that it should have obtained identical results to the original*" is with the authors of the (potential) replication.

3.2 Definition

We propose a straightforward definition: to qualify as a policing replication, the replication should directly challenge a previously published empirical paper and address this original paper prominently, that is, in the title or abstract. The rationale is that an act of policing must be directly attributable to a case. Just like previous papers conceptualizing sub-types of replications, we acknowledge that this is no clear-cut definition. Thus, how does our proposal relate to existing definitions? Taking the *I4R* definitions in Table 1, 'computational replications' and 'robustness replications' are policing replications in virtually all cases. Yet still, if the replicating paper does not directly address the original paper, it is left to the discretion of the reader to update their priors about that original paper – and, hence, we would not call it a policing replication. Also, 'direct replications' and 'conceptual replications', in principle, qualify as policing if they directly challenge the original paper. In many cases, though, they do not make this challenge explicit, for example to claim novelty.[5]

The term policing is meant to convey that empirical scientific discovery needs to be controlled systematically to institutionalize incentives that prevent questionable research practices and fraud.[6] We acknowledge that the term might evoke some negative connotations. We use policing in its very positive sense, that is, a regulatory act preventing intentional or unintentional bad behavior. The police do not sentence. The police only investigate and compile evidence for a case. This evidence is then used

[5] To take the example from footnote four above, if Magnan et al. (2015) directly challenged Conley and Udry (2010) in their abstract, for example because the original paper made very strong claims about generalizability to other contexts and other technologies, Magnan et al. would clearly police the Conley and Udry result.

[6] The need for more *policing* in science was blatantly showcased recently in a high-profile case of alleged fraud in Alzheimer's disease research in neuroscience (Piller, 2022).

by prosecutors and, potentially, a verdict is pronounced by a court of law. In this sense, a policing replication investigates a previously published paper – the role of the prosecutor and the court of law is with the scientific community as the readership. An excellent piece of scholarship in this regard is Ozier (2021).

3.3. Policing replication rates

We now push further by asking how many policing replications are being published. We screened all papers published in the Top 50 economics journals using Scopus (of which we eventually included 42)[7] between 2010 and 2020 for whether they are policing replications (thereby looking for the *total replication rate*). We scraped all papers that

- directly cite another paper in their abstract or title, or
- include the word "comment" in the title, or
- include the word "replic*", "reanal*", or "revisit*" in title, abstract, or keywords.

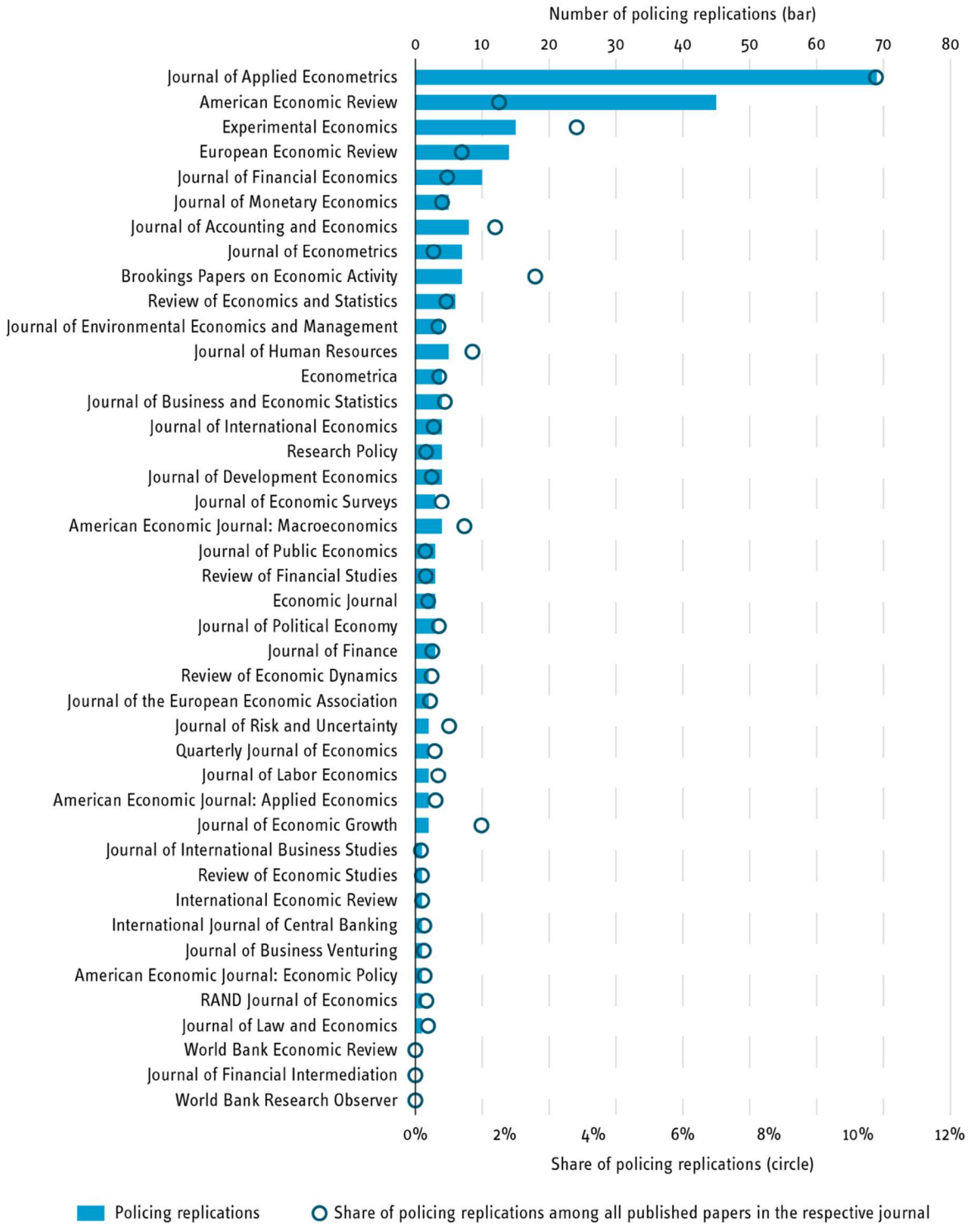
Out of 29,682 papers published in total in the Top 50 journals between 2010 and 2020, 2,787 papers meet these formalized search criteria (see Table A1 in the Appendix for a comprehensive list).[8] We read all abstracts of the resulting papers – which is where the policing ambition must become apparent according to our definition – and coded them as policing replications or not. To be conservative, we coded papers for which we were on the fence as policing. We generally included confirmatory replications but excluded corrigenda, errata, and replies.[9]

[7] We used the Top 50 journals as listed on https://ideas.repec.org/top/top.journals.simple.html, accessed last on July 28, 2021. We excluded eight journals: First, journals of federal reserve banks were excluded because they are not listed on Scopus (*Proceedings, Federal Reserve Bank of Cleveland*, *Proceedings, Federal Reserve Bank of San Francisco*, and *Quarterly Review, Federal Reserve Bank of Minneapolis*). We further excluded the *Journal of Economic Perspectives*, the *Journal of Economic Literature*, the *Annual Review of Economics,* and *Foundations and Trends in Econometrics* because they are review journals and unlikely to publish replications. The *Journal of Economic Theory* is excluded because we focus our work on empirical studies. The *Journal of Business* was discontinued in 2006 and is therefore not part of this table, either. Two journals were jointly listed on rank 21, which is the reason for arriving at 42 journals in total.

[8] The search results as well as the coding of the 2,787 papers that meet the formalized search criteria can be obtained from the authors upon request.

[9] A special case are papers that review multiple papers. A priori, we would consider a replication of a limited number of papers as policing, since it would still uncover paper-specific problems. Yet, we would not consider

Figure 1: Policing replications in the Top 50 economics journals between 2010 and 2020



Notes: For the AER, the calculation of the share of policing replications among all published papers includes the Papers and Proceedings issue. When excluded, the share of policing replications increases from 1.9% to 3.3%. Source: Own data.

systematic reviews and meta-analyses as replications. The demaraction when a replication of a limited number of papers turns into a meta-analysis is not always clear. We tried to be conservative by coding papers that, in the abstract, sound like a review on a small number of papers as policing replication. These decisions should not be consequential, though, as only <0.1% of the papers fulfulling the formalized search criteria are systematic reviews.

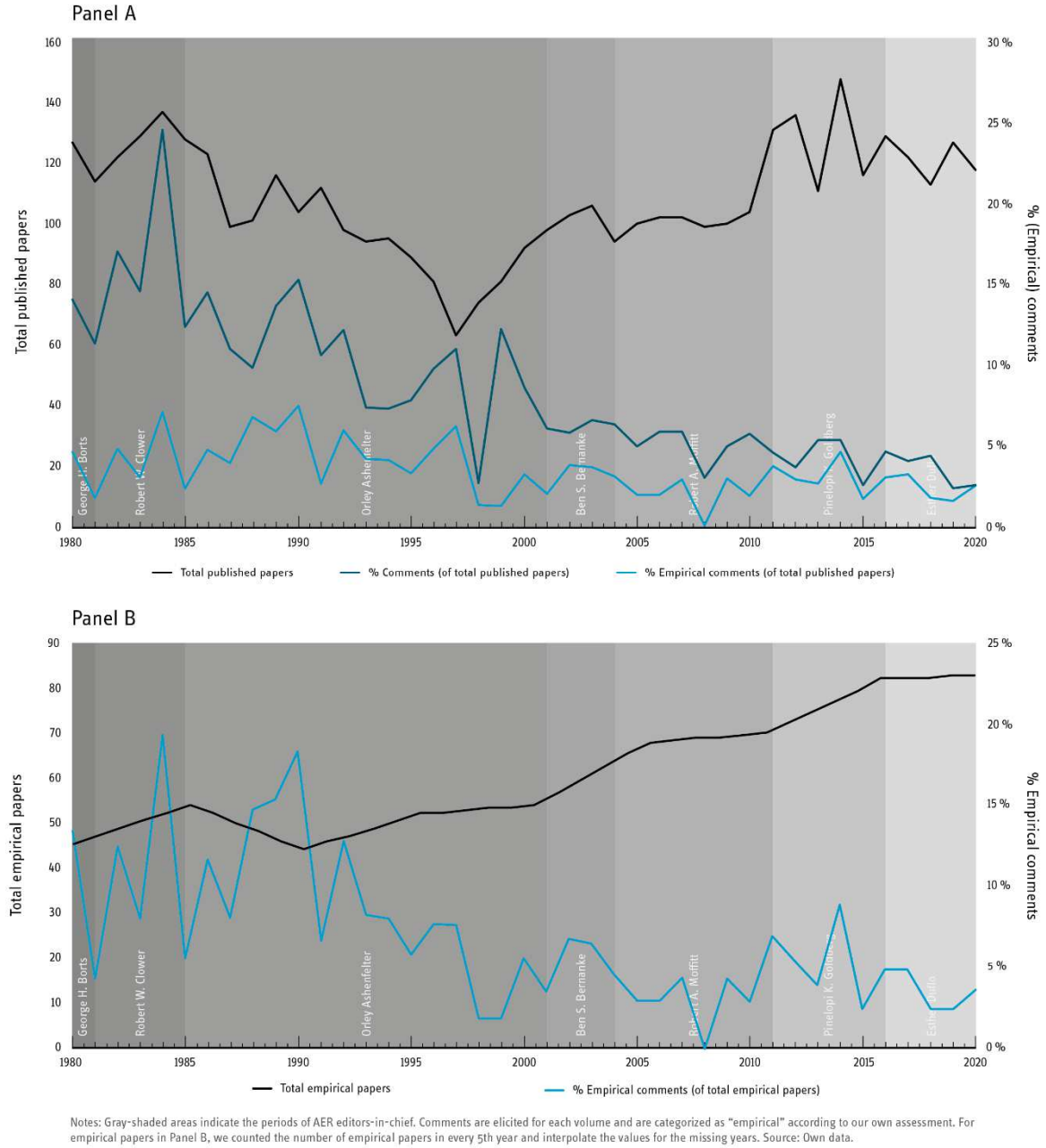
We identified 259 policing replications (i.e., 0.9% of all published papers), of which 238 papers cite the replicated study in the title or abstract.[10] As can be seen in Figure 1, three of the 42 journals have not published any policing replication since 2010 and 30 have published less than five, while two journals (the *AER* and the *Journal of Applied Econometrics*) account for almost 50% of all published policing replications. A priori, we do not know whether the patterns between journals reflect different editorial policies (except for the *Journal of Applied Econometrics (JAE)* that has a dedicated replication section) – or driven by the supply side, this is, authors submitting more policing replications to certain journals. In terms of the economic sub-disciplines, though, we noticed during our coding work that the lion's share of policing replications, also among those in the AER and the JAE, is in macro-, financial, and behavioral economics (including laboratory experiments). We found very few policing replications in applied fields like development, environmental, health and labor economics.

Our review approach has two potential sources of bias: the formalized search criteria might miss papers that we would code as policing and our coding itself is – to some degree – a subjective decision. We therefore corroborate our finding using two different perspectives. First, we isolate those replications identified in Berry et al. (2017), Sukhtankar (2017), and Mueller-Langer et al. (2019) that qualify as policing replications by going through their abstracts and coding them as policing or not. This still involves our coding but uses a different database. For Berry et al. (2017), only one out of the 52 papers they code as replications qualifies as policing. For Sukhtankar (2017), 50 out of 71 papers meet our policing criterion, including 20 replications that have hitherto only been published as working papers. Policing replication rates in these two reviews are, hence, at 1.4% and 2.6%, respectively (see Table A2 in the Appendix). Recall that both reviews look at selective replication rates. In Mueller-

[10] A comprehensive list of all policing replications and their corresponding original papers can be obtained from the authors upon request.

Langer et al. (2019), a review of total replication rates like ours, all except nine papers coded by the authors as replications are policing replications, leading to a policing replication rate of 0.1%.

Figure 2: Papers and comments published in the *AER* between 1980 and 2020



Notes: Gray-shaded areas indicate the periods of AER editors-in-chief. Comments are elicited for each volume and are categorized as "empirical" according to our own assessment. For empirical papers in Panel B, we counted the number of empirical papers in every 5th year and interpolate the values for the missing years. Source: Own data.

Second, we zoom into the *AER*, one of the two journals that publishes the most policing replications, and investigate how many comments it has published since 1980. Comments in the *AER* discuss and challenge papers that were published in the journal

before, hence, arguably have strong policing components. Panel A of Figure 2 shows that there is a continuous downward sloping trend of comments in total, from a high level of between 10 and 20% of all papers in the 80s and 90s to below 5% in the early 2010s and 2-3% in the most recent years.[11] Note that comments in the *AER* can also be theoretical, while our definition of policing replication requires empirical examination. We therefore distinguish between empirical and theoretical comments, showing that indeed in the 80s and 90s many comments were theoretical. The share of empirical comments among all papers has also decreased, but much less steeply. Nevertheless, this drop in empirical comments is noteworthy considering the sharp increase in empirical work in economics, including in the *AER* (Angrist et al., 2017).[12]

As Panel B of Figure 2 shows, the number of empirical papers in total has increased from about 50 between 1980 and 1990 to about 80 today, which is an increase in the percentage of empirical papers considering the number of all published papers has not increased since 1980, as shown in Panel A. Hence, the share of empirical comments in total empirical papers has declined sharply over time. Hamermesh (2017) conjectures that this decline in publishing replications reflects the editors' citation-maximizing strategy, since replications and comments are, on average, cited less than original work, as also pointed out by Ankel-Peters et al. (2022) and Whaples (2006).

4. Publishing replications: Journal policies

Motivated by the results from the previous section, we conducted a simple survey in collaboration with the *I4R* among the editors of the Top 50[13] journals about the journal's policy to publish what we call policing replications (see Figure 4 for the exact wording used in the survey). In total, 33 of 42 contacted editors responded. Figure 3 summarizes the results; the responses disaggregated by journal can be found in Table

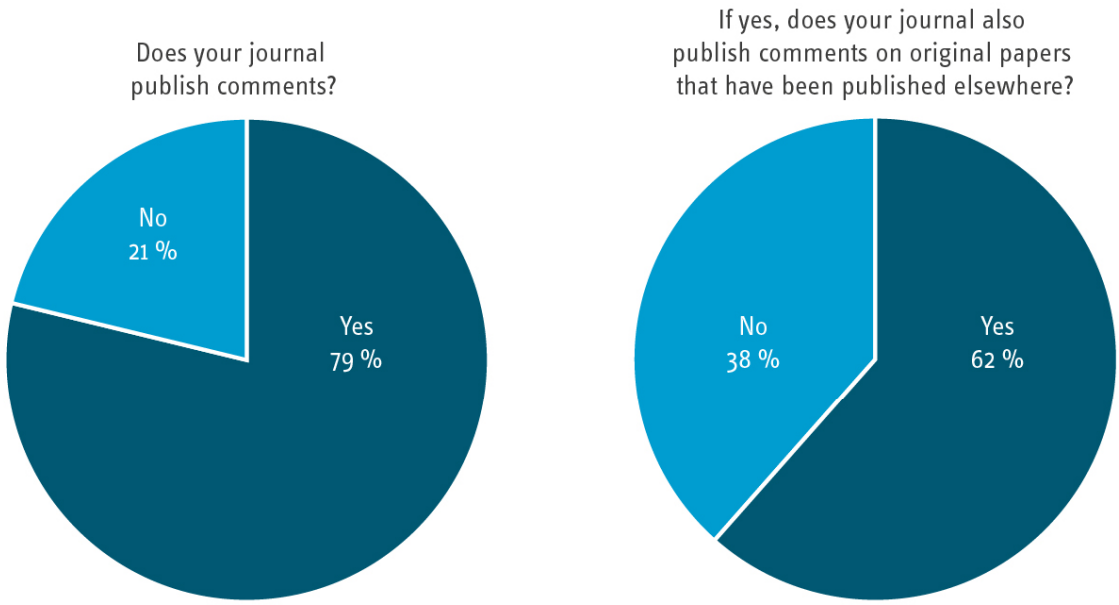
[11] This trend was diagnosed for an earlier period already in Coelho et al. (2005) and has been described in Hamermesh (2017) as well.

[12] Since 2019 the AEA hosts a new journal, the *American Economic Review: Insights* (*AER:I*) and, at the same time, the *AER* stopped featuring a specific short paper section. While one might suspect that comments have been moved from *AER*'s short paper section to *AER:I*, the *AER:I* has not published any comment since its inception.

[13] We covered the same 42 journals out of the Top 50 journals that were also included in our review, see previous section.

A3 in the Appendix. The majority of editors, 79%, stated that their journal publishes policing replications of papers originally published in their own journal. Of those, 62% responded that policing replications of papers in other journals are generally published. Many editors, though, added that the comment would have to make a more general contribution over and above the replication itself (see the detailed responses on the *I4R* website[14]). In addition, we checked the websites of these 42 journals for whether their *Aims & Scope* or *Guide for Authors* state that replications or comments are considered for publication: Seven (17%) do so (see Table A3 in the Appendix.)[15]

Figure 3: Results of the survey among the editors of the Top 50 journals in economics



Notes: We contacted 42 journals, of which 33 responded. 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?

Moreover, it is noteworthy that recently new opportunities to publish replications have emerged. The *Journal of Comments and Replications in Economics* sets out to publish replications defined *"as any study that directly addresses the reliability of a specific claim*

[14] See https://i4replication.org/publishing.html.
[15] We checked the journal websites on December 7, 2022.

from a previously published study" – a notion that corresponds to our definition of policing replication. Several journals, some new and some established, now prominently invite submissions of replications in their *Aims & Scope*, for example the *Journal of Political Economy: Microeconomics,* the *Journal of the Economic Science Association,* and *Q Open*. Others like the *Journal of the European Economic Association* and *Energy Economics* explicitly invite replications in their guide for authors. Not least, initiatives like *BITSS* and *I4R* provide different fora for policing replications, including a newly launched discussion paper series.[16]

While these are good steps in the right direction, attempts of journals in the past to indicate demand for replications have not automatically led to a supply of replications from authors. As Hamermesh (2007) notes, the *Journal of Political Economy* discontinued a section called '*Confirmations and Contradictions*' in 1999, which had mostly published comments using new data, and *Labour Economics* altered their policy which at the beginning emphasized that they welcomed replications; both did this due to a lack of submissions. This is in line with Sukhtankar (2017), who extended his review of published replications to the grey literature: while the policing replication rate increases when working papers are included, it is still only at 4.4%. This suggests a low level of hidden replications that do not make it into journals.

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 policing replications.

5. Conclusion

Complementing the existing definitions of replications and their sub-types, we have proposed a dedicated type of replication that polices previously published work in the spirit of Merton's organized skepticism. We have also found that below 1% of published papers indeed police previous work (the total policing replication rate), and between 1.4% and 2.6% of influential papers have been subject to a policing replication

[16] One of the authors of this paper co-edits the *I4R* discussion paper series.

(the selective policing replication rate). Whether this is a reason for concern or not depends on one's prior about whether there is a replicability problem in economics. We believe some recent meta-scientific work suggests there is.

Note that our own policing replication rates are based on peer-reviewed replications. A lot of replication work is happening in economics classes at both graduate and undergraduate levels where influential papers are re-analysed or subject to robustness checks. Vilhuber (2020a) argues that only a fraction of this work will be published, and this fraction might be biased towards non-confirmative findings. In general, it is likely that, given the current incentive structure and replication publication culture in economics, mostly unsuccessful replications are published, as it was diagnosed for psychology in Bryan et al. (2019). We contend, though, that the sensitive work of policing replications in Merton's spirit should not be left to students who are then supposed to confront powerful original authors with potential problems. Experienced scholars should engage in policing replications as well.

To this end, incentives for researchers are needed. Our claim (based on Clemens, 2017) that authors of replications should assume the burden-of-proof is at odds with the current novelty norm in economics. This novelty norm makes it difficult to publish replications in journals that pay off for academic careers. Moreover, policing replications are often perceived as hostile in the profession. Both of these make it a risky career strategy, especially for young scholars (Hamermesh, 2017; Janz, 2015).

Clearly, reforms and incentives are needed to catalyse a cultural change. Coffman et al. (2017) argue that change must come from the top down, and they call on journals to offer a regular section for replications in each issue (which some do, see the previous section). Not least, "*citations to the original paper [should] include citations to its replication*" (Coffman et al., 2017) – something that is not happening as of today (Ankel-Peters et al., 2022). For psychology, Hardwicke et al. (2021), Schafmeister (2021), and von Hippel (2022) find that replications are mostly neglected in terms of citations and even failed replications do not significantly reduce the citations of original papers. As Whaples (2006) points out, low citation counts also lead to a negative feedback loop because

publishing comments and replications are costly for editors who mostly pursue a citation-maximization strategy (Card and DellaVigna, 2020). In this regard, it is peculiar that most economics journals do not provide links to replications on the replicated paper's website – something that is standard in other professions. The *AER*, for example, is leading in terms of publishing policing replications, as we have shown. However, as of December 2022, it does not provide links on the original paper's website to comments that are published in the *AER* itself. Changing this would be a simple step toward ameliorating the visibility of replications and, thereby, probably also increase their citations.

We also reiterate a very simple and straightforward proposal made by Clemens (2017): The *American Economic Association* and the *Journal of Economic Literature (JEL)* could add explicit categories to the JEL code structure on the different types of replications. This would help to clarify the terminology and at the same time signal that replications are endorsed by the profession's flagship association. Such JEL codes would also facilitate finding replications, hence, including them in systematic reviews and overview articles (Coffman et al., 2017).

Furthermore, leading economics journals should make explicit whether they generally accept or even encourage comments and replications. As we have shown, only 17% of the top 50 journals do this on their websites. But perhaps a cultural transformation really is underway. The verve the AEA data editor has for making data and code available is remarkable (see Vilhuber, 2020b). As our editor survey in Section 4 has indicated, most top journals are, indeed, open to publishing replications. And, not least, *BITSS* and *I4R* facilitate and mainstream replication in the social sciences and, together with other nascent dynamics in economics such as new journals with replication foci, this might ultimately lead to rising replication rates in the coming years.

References

Anderson, M.L., 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association 103, 1481–1495. https://doi.org/10.1198/016214508000000841

Angrist, J., Azoulay, P., Ellison, G., Hill, R., Lu, S.F., 2017. Economic Research Evolves: Fields and Styles. American Economic Review 107, 293–297. https://doi.org/10.1257/aer.p20171117

Ankel-Peters, J., Fiala, N., Neubauer, F., 2022. A Hard Science? Comments in the American Economic Review. Mimeo.

Askarov, Z., Doucouliagos, A., Doucouliagos, H., Stanley, T.D., 2022. The Significance of Data-Sharing Policy. Journal of the European Economic Association jvac053. https://doi.org/10.1093/jeea/jvac053

Berry, J., Coffman, L.C., Hanley, D., Gihleb, R., Wilson, A.J., 2017. Assessing the Rate of Replication in Economics. American Economic Review 107, 27–31. https://doi.org/10.1257/aer.p20171119

Brodeur, A., Cook, N., Heyes, A., 2022a. We Need to Talk about Mechanical Turk: What 22,989 Hypothesis Tests Tell Us about Publication Bias and p-Hacking in Online Experiments. IZA Discussion Papers No. 15478 57.

Brodeur, A., Cook, N., Heyes, A., 2022b. Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics: Reply. American Economic Review 112, 3137–3139. https://doi.org/10.1257/aer.20220277

Brodeur, A., Cook, N., Heyes, A., 2020. Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics. American Economic Review 110, 3634–3660. https://doi.org/10.1257/aer.20190687

Brodeur, A., Cook, N., Neisser, C., 2022c. P-Hacking, Data Type and Data-Sharing Policy. IZA Discussion Papers No. 15586. https://doi.org/10.2139/ssrn.4232716

Brodeur, A., Lé, M., Sangnier, M., Zylberberg, Y., 2016. Star Wars: The Empirics Strike Back. American Economic Journal: Applied Economics 8, 1–32. https://doi.org/10.1257/app.20150044

Bryan, C.J., Yeager, D.S., O'Brien, J.M., 2019. Replicator Degrees of Freedom Allow Publication of Misleading Failures to Replicate. Proceedings of the National Academy of Sciences 116, 25535–25545. https://doi.org/10.1073/pnas.1910951116

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H., 2016. Evaluating Replicability of Laboratory Experiments in Economics. Science 351, 1433–1436. https://doi.org/10.1126/science.aaf0918

Card, D., DellaVigna, S., 2020. What Do Editors Maximize? Evidence from Four Economics Journals. The Review of Economics and Statistics 102, 195–217. https://doi.org/10.1162/rest_a_00839

Christensen, G., Miguel, E., 2018. Transparency, Reproducibility, and the Credibility of Economics Research. Journal of Economic Literature 56, 920–980. https://doi.org/10.1257/jel.20171350

Christensen, G., Wang, Z., Paluck, E.L., Swanson, N., Birke, D.J., Miguel, E., Littman, R., 2020. Open Science Practices are on the Rise: The State of Social Science (3S) Survey. Working Paper. https://doi.org/10.31222/osf.io/5rksu

Clemens, M.A., 2017. The Meaning of Failed Replications: A Review and Proposal. Journal of Economic Surveys 31, 326–342. https://doi.org/10.1111/joes.12139

Coelho, P.R.P., De Worken-Eley III, F., McClure, J.E., 2005. Decline in Critical Commentary, 1963-2004. Econ Journal Watch 2, 355–361.

Coffman, L.C., Niederle, M., Wilson, A.J., 2017. A Proposal to Organize and Promote Replications. American Economic Review 107, 41–45. https://doi.org/10.1257/aer.p20171122

Conley, T.G., Udry, C.R., 2010. Learning about a New Technology: Pineapple in Ghana. American Economic Review 100, 35–69. https://doi.org/10.1257/aer.100.1.35

Dahal, M., Fiala, N., 2020. What Do We Know about the Impact of Microfinance? The Problems of Power and Precision. World Development 128, 104773. https://doi.org/10.1016/j.worlddev.2019.104773

Dreber, A., Johannesson, M., 2022. A Framework for Evaluating Reproducibility and Replicability in the Social Sciences. Mimeo.

Ferraro, P.J., Shukla, P., 2022. Credibility Crisis in Agricultural Economics. Applied Economic Perspectives and Policy 1–17. https://doi.org/10.1002/aepp.13323

Ferraro, P.J., Shukla, P., 2020. Feature—Is a Replicability Crisis on the Horizon for Environmental and Resource Economics? Review of Environmental Economics and Policy 14, 339–351. https://doi.org/10.1093/reep/reaa011

Fink, G., McConnell, M., Vollmer, S., 2014. Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures. Journal of Development Effectiveness 6, 44–57. https://doi.org/10.1080/19439342.2013.875054

Foote, C.L., Goetz, C.F., 2008. The Impact of Legalized Abortion on Crime: Comment. Quarterly Journal of Economics 123, 407–423. https://doi.org/10.1162/qjec.2008.123.1.407

Freese, J., Peterson, D., 2017. Replication in Social Science. Annual Review of Sociology 43, 147–165.

Hamermesh, D.S., 2017. Replication in Labor Economics: Evidence from Data and what it Suggests. American Economic Review 107, 37–40. https://doi.org/10.1257/aer.p20171121

Hamermesh, D.S., 2007. Viewpoint: Replication in Economics. Canadian Journal of Economics/Revue canadienne d'économique 40, 715–733. https://doi.org/10.1111/j.1365-2966.2007.00428.x

Hardwicke, T.E., Szűcs, D., Thibault, R.T., Crüwell, S., van den Akker, O.R., Nuijten, M.B., Ioannidis, J.P.A., 2021. Citation Patterns Following a Strongly Contradictory Replication Result: Four Case Studies From Psychology. Advances in Methods and

Practices in Psychological Science 4, 251524592110408. https://doi.org/10.1177/25152459211040837

Hensel, P.G., 2021. Reproducibility and Replicability Crisis: How Management Compares to Psychology and Economics – A Systematic Review of Literature. European Management Journal 39, 577–594. https://doi.org/10.1016/j.emj.2021.01.002

Höffler, J.H., 2017. Replication and Economics Journal Policies. American Economic Review 107, 52–55. https://doi.org/10.1257/aer.p20171032

Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J.R., Burli, P., Chen, N., Grieco, P., Ekpe, G., Pugatch, T., Saavedra, M., Stopnitzky, Y., 2021. The Influence of Hidden Researcher Decisions in Applied Microeconomics. Economic Inquiry 59, 944–960. https://doi.org/10.1111/ecin.12992

Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H., 2017. The Power of Bias in Economics Research. The Economic Journal 127, F236–F265. https://doi.org/10.1111/ecoj.12461

Janz, N., 2015. Bringing the Gold Standard into the Classroom: Replication in University Teaching. International Studies Perspectives 392–407. https://doi.org/10.1111/insp.12104

John, L.K., Loewenstein, G., Prelec, D., 2012. Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. Psychological Science 23, 524–532. https://doi.org/10.1177/0956797611430953

Kerr, N.L., 1998. HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 2, 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kranz, S., Pütz, P., 2022. Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics: Comment. American Economic Review 112, 3124–3136. https://doi.org/10.1257/aer.20210121

Laitin, D.D., 2013. Fisheries Management. Political Analysis 21, 42–47. https://doi.org/10.1093/pan/mps033

Magnan, N., Spielman, D.J., Lybbert, T.J., Gulati, K., 2015. Leveling with Friends: Social Networks and Indian Farmers' Demand for a Technology with Heterogeneous Benefits. Journal of Development Economics 116, 223–251. https://doi.org/10.1016/j.jdeveco.2015.05.003

Maxwell, S.E., Lau, M.Y., Howard, G.S., 2015. Is Psychology Suffering from a Replication Crisis? What Does "Failure to Replicate" Really Mean? American Psychologist 70, 487–498. https://doi.org/10.1037/a0039400

Merton, R.K., 1973. The Normative Structure of Science, in: Storer, N.W. (Ed.), The Sociology of Science. The University of Chicago Press, Chicago.

Miguel, E., 2021. Evidence on Research Transparency in Economics. Journal of Economic Perspectives 35, 193–214. https://doi.org/10.1257/jep.35.3.193

Mueller-Langer, F., Fecher, B., Harhoff, D., Wagner, G.G., 2019. Replication Studies in Economics—How Many and Which Papers Are Chosen for Replication, and Why? Research Policy 48, 62–83. https://doi.org/10.1016/j.respol.2018.07.019

Ofosu, G.K., Posner, D.N., 2021. Pre-Analysis Plans: An Early Stocktaking. Perspectives on Politics 1–17. https://doi.org/10.1017/S1537592721000931

O'Grady, C., 2020. Misconduct Allegations Push Psychology Hero off his Pedestal. Science. https://doi.org/10.1126/science.abd8375

Open Science Collaboration, 2015. Estimating the Reproducibility of Psychological Science. Science 349, aac4716. https://doi.org/10.1126/science.aac4716

Ozier, O., 2021. Replication Redux: The Reproducibility Crisis and the Case of Deworming. The World Bank Research Observer 36, 101–130. https://doi.org/10.1093/wbro/lkaa005

Peters, J., Langbein, J., Roberts, G., 2018. Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity. The World Bank Research Observer 33, 34–64. https://doi.org/10.1093/wbro/lkx005

Piller, C., 2022. Blots on a Field? Science 377, 358–363. https://doi.org/10.1126/science.add9993

Schafmeister, F., 2021. The Effect of Replications on Citation Patterns: Evidence From a Large-Scale Reproducibility Project. Psychological Science 32, 1537–1548. https://doi.org/10.1177/09567976211005767

Servick, K., 2018. Cornell Nutrition Scientist Resigns After Retractions and Research Misconduct Finding. Science. https://doi.org/10.1126/science.aav4979

Simonsohn, U., 2013. Just Post It: The Lesson From Two Cases of Fabricated Data Detected by Statistics Alone. Psychological Science 24, 1875–1888. https://doi.org/10.1177/0956797613480366

Sukhtankar, S., 2017. Replications in Development Economics. American Economic Review 107, 32–36. https://doi.org/10.1257/aer.p20171120

Vilhuber, L., 2020a. Reproducibility and Replicability in Economics. Harvard Data Science Review 2. https://doi.org/10.1162/99608f92.4f6b9e67

Vilhuber, L., 2020b. AEA Data and Code Availability Policy. AEA Papers and Proceedings 110, 776–778. https://doi.org/10.1257/pandp.110.776

Vivalt, E., 2020. How Much Can We Generalize From Impact Evaluations? Journal of the European Economic Association 18, 3045–3089. https://doi.org/10.1093/jeea/jvaa019

von Hippel, P.T., 2022. Is Psychological Science Self-Correcting? Citations Before and After Successful and Failed Replications. Perspectives on Psychological Science 174569162110725. https://doi.org/10.1177/17456916211072525

Whaples, R., 2006. The Costs of Critical Commentary in Economics Journals. Econ Journal Watch 3, 275–282.

Appendix

Table A1: Total number of published papers and policing replications in the Top 50 economics journals between 2010 and 2020

Journal	Total number of published papers	Papers meeting formalized search criteria	Number (and share*) of policing replications	Number (and share*) of policing replications that cite the original study in the title or abstract
Quarterly Journal of Economics	446	20	2 (0.4%)	2 (0.4%)
Econometrica	727	141	4 (0.6%)	4 (0.6%)
Journal of Economic Growth	135	6	2 (1.5%)	2 (1.5%)
Journal of Financial Economics	1,387	94	10 (0.7%)	10 (0.7%)
Review of Financial Studies	1,194	66	3 (0.3%)	3 (0.3%)
American Economic Review	2,387	152	45 (1.9%)	45 (1.9%)
Journal of Political Economy	565	32	3 (0.5%)	2 (0.4%)
Journal of Finance	774	32	3 (0.4%)	3 (0.4%)
Review of Economic Studies	607	50	1 (0.2%)	1 (0.2%)
Journal of Monetary Economics	845	189	5 (0.6%)	5 (0.6%)
Journal of Labor Economics	380	10	2 (0.5%)	2 (0.5%)
Brookings Papers on Economic Activity	260	134	7 (2.7%)	0 (0%)
Journal of Accounting and Economics	444	83	8 (1.8%)	8 (1.8%)
American Economic Journal: Macroeconomics	361	28	4 (1.1%)	4 (1.1%)
Journal of the European Economic Association	579	39	2 (0.3%)	1 (0.2%)
Journal of Econometrics	1,685	398	7 (0.4%)	7 (0.4%)
American Economic Journal: Applied Economics	445	7	2 (0.4%)	2 (0.4%)
RAND Journal of Economics	393	12	1 (0.3%)	0 (0%)
Review of Economics and Statistics	852	66	6 (0.7%)	6 (0.7%)
Journal of Applied Econometrics	668	126	69 (10.3%)	65 (9.7%)
Economic Journal	980	63	3 (0.3%)	3 (0.3%)
Journal of International Economics	947	116	4 (0.4%)	3 (0.3%)
Journal of Financial Intermediation	316	11	0 (0%)	0 (0%)
Journal of Business Venturing	508	14	1 (0.2%)	1 (0.2%)
Journal of Business and Economic Statistics	588	113	4 (0.7%)	2 (0.3%)
Journal of Public Economics	1,273	69	3 (0.2%)	2 (0.2%)
World Bank Economic Review	364	14	0 (0%)	0 (0%)
Journal of International Business Studies	716	78	1 (0.1%)	1 (0.1%)
Journal of Development Economics	1,058	51	4 (0.4%)	3 (0.3%)
Experimental Economics	416	61	15 (3.6%)	15 (3.6%)
American Economic Journal: Economic Policy	456	13	1 (0.2%)	1 (0.2%)
Journal of Environmental Economics and Management	741	41	4 (0.5%)	4 (0.5%)
Journal of Law and Economics	335	7	1 (0.3%)	1 (0.3%)
Journal of Human Resources	388	16	5 (1.3%)	5 (1.3%)
World Bank Research Observer	97	5	0 (0%)	0 (0%)
Journal of Risk and Uncertainty	260	30	2 (0.8%)	1 (0.4%)

Research Policy	1,607	90	4 (0.2%)	3 (0.2%)
Journal of Economic Surveys	508	42	3 (0.6%)	3 (0.6%)
International Economic Review	610	37	1 (0.2%)	1 (0.2%)
International Journal of Central Banking	484	37	1 (0.2%)	1 (0.2%)
European Economic Review	1,346	111	14 (1%)	14 (1%)
Review of Economic Dynamics	550	83	2 (0.4%)	2 (0.4%)
Total	29,682	2,787	259 (0.9%)	238 (0.8%)

Notes: We use the Top 50 journals as listed on https://ideas.repec.org/top/top.journals.simple.html, accessed on July 28, 2021. We exclude journals of federal reserve banks because they are not listed on Scopus ("*Proceedings, Federal Reserve Bank of Cleveland*", "*Proceedings, Federal Reserve Bank of San Francisco*", and "*Quarterly Review, Federal Reserve Bank of Minneapolis*"). We further exclude the *Journal of Economic Perspectives*, the *Journal of Economic Literature*, the *Annual Review of Economics*, and *Foundations and Trends in Econometrics* because they are review journals and unlikely to publish replication studies. The *Journal of Economic Theory* is excluded because we concentrate our work on empirical studies. The *Journal of Business* was discontinued in 2006 and is therefore not part of this table, either. Two journals were jointly listed on rank 21, which is the reason for arriving at 42 journals in total. *Share of policing replications in the total number of published papers.

Table A2: Policing replications in Berry et al. (2017), Sukhtankar (2017), and Mueller-Langer et al. (2019)

Paper	Number of observations in sample (1)	Sample (2)	Number of replications/replicated papers[1] as coded in respective paper (3)	Number of policing replications (our coding) (4)	Share of policing replications in all replications [(4)/(3)] (5)	Replication rate according to paper [(3)/(1)] (6)	Policing replication rate based on our coding [(4)/(1)] (7)
Berry et al. (2017)	70	All papers of *AER* volume 100, published in 2010	42	1[#]	2%	60.0%*	1.4%
Sukhtankar (2017)	1,138	Empirical papers with "O" classification in 10 journals during 2000-2015	71	30[#]	42%	6.2%[+]	2.6%
Mueller-Langer et al. (2019)	126,505	All articles published in the Top 50 economics journals during 1974 – 2014	130	121	93%	0.1%	0.1%

Notes: [1]Berry et al (2007) and Sukhtankar (2017) examine the selective replication rate, which estimates how many published papers are replicated, while Mueller-Langer et al. (2019) examines the total replication rate, which estimates how many papers are replications. [#]Berry et al (2007) have coded 52 citing papers as replications of 42 original papers. The one paper we coded as a policing replication only referred to one original paper. Similarly, the 30 papers coded as policing replications in Sukhtankar (2017) replicate one original paper each. *The reported replication rate from Berry et al. (2017) presented here is the one in which the authors include any of the three categories branded by them as "Replication", "Extension", or "Robustness". [+]The reported replication rate from Sukhtankar (2017) is based on a review of both peer-reviewed replications and working papers. Our coding of policing replications in column (4) and the calculation of the policing replication rate in column (7) only considers replications from peer-reviewed journals.

Table A3: Responses to the survey among editors of the Top 50 economics journals

Journal	Editor[1,2] (responding editors underlined)	[Q1] Do you publish comments[3]?	[Q2] If yes, do you also publish comments on original papers that have been published elsewhere?	Does the website mention that the journal publishes comments or replications?[4]
American Economic Journal: Economic Policy	Erzo Luttmer	Yes	Yes	No
American Economic Journal: Macroeconomics	Simon Gilchrist	Yes	No	No
American Economic Review	Esther Duflo	Yes	Yes	Yes
Brookings Papers on Economic Activity	Janice Eberly, James Stock	No	No	Yes
Econometrica	Guido Imbens	Yes	Yes	No
Economic Journal	Francesco Lippi	Yes	Yes	No
Experimental Economics	John Duffy, Lata Gangadharan, Ragan Petrie, Arno Riedl, Roberto Weber	Yes	Yes	Yes
International Economic Review	Dirk Krueger	Yes	No	No
International Journal of Central Banking	Luc Laeven	Yes	No	No
Journal of Accounting and Economics	John Core, Wayne Guay, Michelle Hanlon, Robert Holthausen, Mark Lang, Joanna Wu	Yes	No	No
Journal of Applied Econometrics	Barbara Rossi	No	No	Yes
Journal of Business & Economic Statistics	Jianqing Fan, Christian Hansen	Yes	Yes	No
Journal of Development Economics	Andrew Foster	Yes	Yes	No
Journal of Economic Growth	Oded Galor	No	No	No
Journal of Economic Surveys	Iris Claus, Leslie Oxley	Yes	No	No
Journal of Environmental Economics and Management	Roger von Haefen, Andreas Lange, Molly Lipscomb	Yes	Yes	Yes
Journal of Finance	Stefan Nagel	Yes	No	No
Journal of Financial Economics	Toni Whited	Yes	Yes	No
Journal of Human Resources	Anna Aizer	No	No	No
Journal of International Economics	Costas Arkolakis, Martin Uribe	Yes	Yes	No
Journal of Labor Economics	Kevin Lang	Yes	Yes	No
Journal of Law and Economics	Dennis Carlton, Dhammika Dharmapala, Richard Holden, Nathan Miller, Sam Peltzman, and Christopher Snyder	Yes	No	No
Journal of Political Economy	Magne Mogstad	Yes	Yes	Yes
Journal of Public Economics	Wojciech Kopczuk and Nathaniel Hendren	Yes	No	No
Journal of Risk and Uncertainty	Kip Viscusi	No	No	No

Journal of the European Economic Association	Imran Rasul	Yes	Yes	Yes
RAND Journal of Economics	Kathleen Mullen	Not yet	No	No
Research Policy	Ben Martin, Maria Savona, Anna Bergek, Alex Coad, Maryann Feldman, Elisa Giuliani, Adam Jaffe, Martin Kenney, Keun Lee, Kazuyuki Motohashi, Paul Nightingale, Ammon Salter, John Walsh	Yes	No	No
Review of Economic Dynamics	Loukas Karabarbounis	Yes	Yes	No
Review of Economic Studies	Nicola Fuchs-Schündeln	Yes	No	No
Review of Economics and Statistics	Will Dobbie	Yes	Yes	No
Review of Financial Studies	Itay Goldstein	No	No	No
World Bank Research Observer	Peter Lanjouw	Yes	Yes	No
No responses from editors				
American Economic Journal: Applied Economics				No
European Economic Review				No
Journal of Business Venturing				No
Journal of Econometrics				No
Journal of Financial Intermediation				No
Journal of International Business Studies				No
Journal of Monetary Economics				No
Quarterly Journal of Economics				No
World Bank Economic Review				No

Notes: [1]The following editors were replaced after we conducted our survey but before publication: Stefan Nagel at the *Journal of Finance* (replaced by Antoinette Schoar), Kathleen Mullen at the *RAND Journal of Economics* (replaced by Craig Bond), Luc Laeven at the *International Journal of Central Banking* (replaced by Christopher Waller), Lata Gangadharan at *Experimental Economics* (replaced by Andreas Ortmann), Robert Holthausen at the *Journal of Accounting and Economics* (replaced by E. deHaan), Jianqing Fan and Christian Hansen at the *Journal of Business & Economic Statistics* (replaced by Ivan Canay and Atsushi Inoue), Florin Bilbiie at the *European Economic Review* (replaced by Evi Pappa), Urban Jermann at the *Journal of Monetary Economics* (replaced by Boragan Aruoba). [2]We did not get responses from the main editors, but the administrative associates or managing editors. The website search was done on December 5, 2022. [3]In the email to the editors, we specified that "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". [4]We scanned the journal websites ourselves, while Q1 and Q2 were part of the editor survey.