ANNUAL REVIEWS

Annual Review of Law and Social Science

Replicability in Empirical Legal Research

Jason M. Chin[1],* and Kathryn Zeiler[2],*

[1] School of Law, University of Sydney, Sydney, New South Wales 2006, Australia;
email: jason.chin@sydney.edu.au

[2] School of Law, Boston University, Boston, Massachusetts 02215, USA; email: kzeiler@bu.edu

*These authors contributed equally to this article

ANNUAL REVIEWS **CONNECT**

www.annualreviews.org

• Download figures
• Navigate cited references
• Keyword search
• Explore related articles
• Share via email or social media

Keywords

empirical legal research, metascience, replicability, reproducibility,
verification, open science, law and economics, law and psychology

Abstract

As part of a broader methodological reform movement, scientists are
increasingly interested in improving the replicability of their research.
Replicability allows others to perform replications to explore potential
errors and statistical issues that might call the original results into question.
Little attention, however, has been paid to the state of replicability in the
field of empirical legal research (ELR). Quality is especially important in
this field because empirical legal researchers produce work that is regularly
relied upon by courts and other legal bodies. In this review, we summarize
the current state of ELR relative to the broader movement toward repli-
cability in the social sciences. As part of that aim, we summarize recent
collective replication efforts in ELR and transparency and replicability
guidelines adopted by journals that publish ELR. Based on this review, ELR
seems to be lagging other fields in implementing reforms. We conclude
with suggestions for reforms that might encourage improved replicability.

1. INTRODUCTION

The replication crisis that has struck all scientific fields, from medicine to economics to psychology and beyond, is by now well known (Munafò et al. 2017). Robust evidence has brought to light the toll on scientific progress imposed by questionable research and publication practices, including selective reporting of methods and results (Simmons et al. 2011), use of unverifiable data and methods (Hardwicke et al. 2020a, p. 15), post-hoc conjecturing (Freese & Peterson 2017, p. 155), insufficient sample sizes (Ioannidis 2005), and a bias toward publishing only statistically significant results (Kvarven et al. 2020). The crisis has sparked several reform efforts that seem to slowly be taking root despite ongoing controversies around optimal approaches (Hardwicke et al. 2020a).

One such reform effort focuses on ensuring that researchers have what they need to attempt to replicate published empirical results. We characterize research as replicable if researchers can access all information necessary to attempt a replication (see the sidebar titled Replication Defined). This requires authors to make available the "input data [and steps used to collect it], computational steps, methods, and code, and conditions of analysis" (Natl. Acad. Sci. Eng. Med. 2019,

REPLICATION DEFINED

Replication differs depending on the nature of the data. For studies employing data not generated by experiments, sometimes referred to in ELR as observational studies, replication entails using the same data (or a different sample drawn from the same population) and same methods to verify (i.e., computationally replicate) the reported results (Clemens 2015). This method is akin to auditing the original work to detect coding mistakes, computational and clerical errors, statistical problems related to insufficient power, and other such issues that might call into question the conclusions drawn from the original results (Vazire & Holcombe 2020, pp. 11–12). For studies employing data generated by experimental studies, the standard replication method is to employ the same experiment protocols used in the original study to produce a different, and often larger, data set drawn from a similar population to determine whether the original results can be replicated (see Klein et al. 2014). Failure to replicate is usually attributed to insufficient power, publication bias, imprecise measurements, sampling error, and undisclosed researcher degrees of freedom (Munafò et al. 2017).

Researchers, of course, also make use of publicly available original data and analysis materials to extend the original results (using the same methods and data drawn from a different population) or reanalyze them (using different, perhaps more up-to-date, methods to analyze the same data or a sample drawn from the same population) (Clemens 2015). If these steps are taken, however, without first checking for coding errors, computational errors, and other basic mistakes, we risk compounding the errors (e.g., reusing faulty data) and spending time and resources attempting to explore and explain divergent results that arise not from different data or methods but from undetected mistakes made in the original research. Our focus here is on reforms aimed at ensuring the availability of materials necessary to perform replications and whether researchers are using the information to produce and publish replication studies. Unless we first verify original results, we risk heading down an unproductive path to explore a phenomenon that might not in fact exist. For this reason, we respectfully disagree with Hubbard & Hyman (2019, p. 2), who potentially muddy the waters by classifying extensions (different methods and different data) and refinements (different methods and same data) as replications ("we present a simple framework for classifying replication efforts into four discrete boxes," including extension and refinement). They do, however, helpfully advise "replicators" to reproduce the target study's original results whenever possible.

p. 36).[1] Easy access to the original data, collection methods, and all analysis materials allows others to attempt to produce new results that either confirm or call into question the original study's results and conclusions.

Replicability is central to the research process because it allows researchers to correct errors before they become generally accepted by other researchers and the public. One way that replicability promotes correction is by making replication research possible. In particular, conclusions drawn from results reported in quantitative empirical studies are more trustworthy if they are replicated by other scientists using the original methods and the original or new data drawn from the same or a similar population.[2] A strong commitment to replication also encourages authors of original studies to precisely state a complete list of sufficient and necessary conditions thought to lead to the original results (Howe & Perfors 2018). This leads to more useful theory in the sense that the conditions under which the theory might successfully predict behavior are clearer. Replicability also helps catch more mundane errors, such as statistical mistakes and coding errors (Vazire 2019). Finally, requiring research to be replicable encourages careful work (Mitchell 2004, Munafò et al. 2017) and deters fraud (Miyakawa 2020).

In this review, we explore the state of replicability in the field of empirical legal research (ELR). We define ELR narrowly as quantitative data-driven investigations of law, legal systems, and other legal phenomena.[3] Scholars have been producing quantitative ELR for decades, and the field continues to grow with the increasing availability of relevant data and technological tools and the continuing interest among journal editors (Diamond & Mueller 2010). All justifications supporting efforts to encourage replicable research resonate forcefully in the context of ELR (Donohue 2015). ELR specifically is designed to address legal issues and is cited by experts testifying in front of legislative bodies, legislatures, authors of briefs filed with courts, judges, authors of comments to administrative agencies, and rule makers housed in agencies.[4] Indeed, serious concerns about the replicability of ELR have been raised, with little or no discernible impact (see Epstein & King 2002b, Zeiler 2016).

Our review proceeds as follows. In Section 2, we establish a baseline by describing the replicability movement that has been underway for some time in the social sciences. We summarize

---

[1]More specifically, policies that foster replication include requiring public availability of all data employed in the study along with

> a clear description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study; a clear description of the analysis of data and decisions for exclusion of some data and inclusion of other; for results that depend on statistical inference, a description of the analytic decisions and when these decisions were made and whether the study is exploratory or confirmatory; a discussion of the expected constraints on generality, such as which methodological features the authors think could be varied without affecting the result and which must remain constant; reporting of precision or statistical power; and a discussion of the uncertainty of the measurements, results, and inferences. (Natl. Acad. Sci. Eng. Med. 2019, pp. 175–76)

[2]Generally, replications of observational studies employ the original methods and data, although in some cases a new sample is drawn from the same population. Replications of experimental studies draw a new sample from a similar population using the same experiment design. Replications that increase sample sizes by drawing observations from the same or a similar population can help produce more precise estimates of effect sizes (Zwaan et al. 2018).

[3]The importance of replicability of qualitative case studies in law has been covered elsewhere (e.g., Argyrou 2017).

[4]Zeiler (2016) informally estimates the frequency of citation in judicial opinions of ELR published in student-edited law journals, finding that reliance on such research is common. ELR published in both student- and faculty-edited journals likely is also regularly cited in other types of legal documents.

efforts by journal editors to reform transparency and openness guidelines. We also describe the notoriously problematic results of systematic replication attempts. Although much work remains to be done, efforts in the social sciences are a good baseline against which to compare progress in law. In Section 3, we describe where things stand on the replicability front in law. We summarize the calls for reform, student- and faculty-edited law journal policies related to replicability, and results from the first collective effort (to our knowledge) to encourage ELR replications. Although progress is underway, the picture is bleak relative to other social science fields. In Section 4, we describe several distinctive institutional barriers that stand in the way of law following in the footsteps of social science. In light of these hurdles, we provide a summary of suggestions others have forwarded to move ELR in the direction of fields that have progressed the most on the replicability front. We offer additional ideas. Section 5 concludes with a road map for further study.

2. REPLICABILITY AND REPLICATION IN SOCIAL SCIENCE

This section describes the movement toward replicability in social science (Christensen et al. 2019b) and a meta-research agenda that seeks to measure and guide that movement (Hardwicke et al. 2020a). We discuss the prevalence of replicable research practices in social science, the consequences of poor replicability, initiatives seeking to improve replicability, and early assessments of changes in the field.

2.1. Irreplicable Research and Failed Replications

The growing field of meta-research is finding that much experimental social scientific research is not replicable (Hardwicke et al. 2020b). For example, a recent study examined a random sample of social scientific experimental studies published from 2014 to 2017, assessing, in part, availability of data, materials, and analysis scripts (Hardwicke et al. 2020b), along with "precise descriptions of the data collection process" (Gelman 2017, p. 38). The results suggest a low level of information availability and replicability, finding, for instance, only 11% of studies using publicly available data and just 2% making their analysis scripts available. Other studies and reviews have found similar deficiencies in replicability in social science (Freese & Peterson 2017, Hardwicke et al. 2018, Vanpaemel et al. 2015) and beyond (Vines et al. 2014).

Further evidence for irreplicability in experimental social science can be found in surveys of researchers in a variety of fields (e.g., psychology, economics, ecology, education) using a variety of methods (e.g., experimental, qualitative) about their practices (sometimes called questionable research practices, or *p*-hacking; Fraser et al. 2018, John et al. 2012, MacCoun 2021, Makel et al. 2021, Necker 2014). These practices often involve failing to report key methodological details, such as why outliers were excluded, how the researchers decided to stop data collection, and conditions and measurements that did not work as predicted. Surveys find that such underreporting is common in psychology (John et al. 2012), economics (Necker 2014), and education research (Makel et al. 2021). There is also evidence for questionable research practices in political science (Franco et al. 2015) and business research (Butler et al. 2017).

Irreplicable practices employed in experimental work likely contribute to the surprisingly high number of false and inflated discoveries in the published literature (Beerdsen 2020, Simmons et al. 2011). Notably, the results of many large-scale, preregistered, multi-lab efforts (**Table 1**) have failed to replicate substantial numbers of published social science experiments (see the sidebar titled Preregistration) (Camerer et al. 2016, 2018; Ebersole et al. 2016; Klein et al. 2018b; Open Sci. Collab. 2015). A recent study compared the multi-lab replication results to results reported in meta-analyses designed to produce a single estimate by combining evidence from multiple studies of the same effect (Kvarven et al. 2020). Notably, the meta-analyses used smaller samples and were

Table 1 Results of six large multi-lab replication projects of experiments in social science

Project	Field	No. replication studies	Statistically significant in the same direction as the original
Estimating the replicability of psychological science	Psychology	97	36%
Evaluating replicability of laboratory experiments in economics	Economics	18	61%
Investigating variation in replicability (Many Labs 1)	Psychology	16	88%
Many Labs 2: investigating variation in replicability across samples and settings	Psychology	28	54%
Many Labs 3: Evaluating participant pool quality across the academic semester via replication	Psychology	9	33%
Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015	Social sciences	21	62%

not preregistered. The comparison revealed that effect sizes reported in the meta-analyses were three times greater than those reported in the registered replications. This suggests that greater analytic flexibility—enabled by a lack of replicability—leads to less trustworthy results. Although these projects have focused on replicating experiments, irreplicable practices also threaten the conclusions that can be drawn from observational studies (e.g., researchers adding covariates to their models until they find the relationship they are looking for without disclosing this in the published study; see Bruns & Ioannidis 2016).

Irreplicability also makes detecting miscalculations and other mistakes in published studies more difficult. Without access to raw data and analytic code, it is hard to know whether published summary statistics are correct. This is problematic because errors do appear to be common in social scientific experimental research, and they are often used to draw faulty conclusions. For example, one study in psychology found that half of a sample of articles included a minor error, and one of every eight included an error that changed reported inferences related to statistical significance (Nuijten et al. 2016).

Meta-researchers are revealing a similarly dire situation in nonexperimental social science fields. Problems in the field of econometrics, for example, came to light much earlier than in experimental fields, garnering attention after Learner (1983) issued his well-known critique. Christensen & Miguel (2018) summarize evidence of widespread replicability problems in economics. Many have reported difficulty obtaining materials necessary for replication even when journals publishing the work have adopted explicit disclosure policies (Dewald et al. 1986, Glandon 2010, McCullough 2009). The *American Economic Review*'s Annual Report of the Editors shows a

PREREGISTRATION

Preregistration is public precommitment to a data collection and analysis plan. It promotes transparency and dissuades questionable research practices by recording the researcher's methods prior to their seeing the data (e.g., their preregistered plan for excluding outliers can be compared to how they actually excluded outliers). It also counteracts publication bias (i.e., null findings are less likely to be published, despite providing important information) by creating a public record of the study. Preregistrations can be embargoed until the study is ready for publication. It is required by law in many jurisdictions for clinical medical trials, reflecting the importance of full reporting in that context. See Nosek et al. (2018) for further information. Bavli (2022) discusses the limitations of preregistration and proposes an alternative.

concerning trend related to papers for which the journal waives data disclosure requirements—increasing steadily from 6% in 2005 to 46% in 2016 (Christensen & Miguel (2018).[5]

Recent research suggests that results reported in econometric studies regularly fail to replicate. Chang & Li (2021) attempted to computationally replicate results from 67 macroeconomics articles. Six were excluded owing to the use of proprietary data. They successfully replicated results in 48% of the 61 studies. Missing data accounted for the largest portion of failures. Clemens (2015) lists 13 studies that failed to replicate published results. The uncovered problems included programming errors, errors in data sets, and inability to produce reported results using the original data and analysis code.

2.2. Responses to Irreplicability

The above insights into the troubling state of replicability in both experimental and nonexperimental social science prompted actions and reforms aimed at improving research openness and transparency. These include development of organizations aimed at providing the knowledge, training, and infrastructure required for conducting replicable research, moves to promote and adopt practices designed to increase replicability, and the creation of new academic journals focused more on methods than results (Vazire 2018).

Beginning with organizations, the Berkeley Initiative for Transparency in the Social Sciences aims to "improve the credibility of science by advancing transparency, reproducibility, rigor, and ethics in research" (**https://www.bitss.org/**). It does this by, among other things, sponsoring workshops on research methodology and funding meta-research (Natl. Acad. Sci. Eng. Med. 2018, p. 153). Similarly, the Center for Open Science (COS) strives to "increase openness, integrity, and reproducibility of scholarly research" (Natl. Acad. Sci. Eng. Med. 2018, p. 114). It has led several initiatives, including development of journal guidelines that promote replicable research. It also developed and maintains the Open Science Framework, a free resource for researchers to store their data and study materials, collaborate, and document research workflows and registrations.

New institutions and organizations have helped to move forward initiatives aimed at improving replicability. Notably, in 2014, a group of researchers, funding agencies, and journal editors created the Transparency and Openness Promotion (TOP) Guidelines, which are "[a]uthor guidelines for journals" that "help to promote transparency, openness, and reproducibility" (Nosek et al. 2015). More than 5,000 journals have signed a statement agreeing to consider adopting TOP to some degree, and more than 1,000 have gone on to do (Cent. Open Sci. 2020b). Journals can implement eight original TOP guidelines (e.g., data transparency, study preregistration) at three levels of rigor (e.g., for open data, they range from requiring a statement about whether data are available to requiring that data actually be posted to a trusted repository; see Nosek et al. 2015).

More recently, the COS developed the TOP Factor, a website and database that display how well journal policies accord with the eight TOP guidelines and two new standards (for a total of ten). The first of those new standards is whether the journal has policies to counter publication

---

[5]The *American Economic Review* is the top economics journal. The journal's policy related to proprietary data is included in its Data and Code Availability Policy ["When possible, a private (not to be published) version of the data should be provided to the AEA Data Editor and/or a designated third-party replicator who can provide a third-party reproducibility report" (**https://www.aeaweb.org/journals/data/data-code-policy**)].

bias.[6] This standard considers whether journals accept registered reports (Chambers 2019), a new journal format in which authors submit their methods for peer review prior to collecting and analyzing data. The second new standard is whether the journal offers open science badges (i.e., optional visible symbols added to articles that meet some criterion, such as posting data or materials online; see Kidwell et al. 2016, Rowhani-Farid & Barnett 2018).

Beyond these general guidelines, more targeted initiatives are being developed to improve replicability. For instance, one group recently used a consensus-based procedure to create a reporting checklist for authors of behavioral and social scientific studies (Aczel et al. 2020). That checklist is geared toward experiments, but improved reporting of observational studies has been discussed widely in economics (see Christensen & Miguel 2018, pp. 959–60). In addition, the American Economic Association recognized that, whereas trial registries (public repositories for preregistrations; see clinicaltrials.gov) are common in medical science, registries are lacking for social science. That group then developed a registry for randomized controlled trials (Am. Econ. Assoc. 2020). The COS hosts a more general registry (**https://osf.io/registries**). Together, these initiatives encourage replicability by providing tools and infrastructure (e.g., in the American Economic Association registry) and encouraging the use of advances in checklists and other reporting guidelines. Reporting checklists are correlated with more thorough methodological disclosure, at least in some fields (Han et al. 2017, Moher et al. 2009).

Experimental psychology has also been proactive in developing outlets for research that focus on improving methodology and methodological reporting (*Collabra: Psychology* 2020, see also Simons 2018). For instance, *Collabra: Psychology* (2020) describes its acceptance guidelines as divorced from the actual result and focused instead on the methodology employed:

> The primary criteria for evaluating manuscripts submitted to *Collabra: Psychology* are scientific, methodological, and ethical rigor. We have a high bar for rigor and integrity, and expect authors to submit high quality work and carefully calibrate their claims to the quality of the evidence. However, novelty and potential impact are not considered important criteria in our evaluations - all research that is rigorous and addresses an important question (even if narrow, or purely descriptive, or purely methodological) will be considered, and evaluated regardless of the results.

Economics has also seen several developments to aid the production of replicable research (Christensen & Miguel 2018). Gentzkow & Shapiro (2014) developed a manual to help researchers produce replicable results. Christensen et al. (2019a) published a text on how to do open science, including steps required to ensure replicability. Koenker & Zeileis (2009) discuss methods of "literate programming," which can be used to combine into a single file all statistical analysis along with the paper draft. Racine (2019) explains in detail how to produce replicable econometrics in R, a programming language popular among economists.

At least one social science journal performs prepublication verification. The *American Journal of Political Science* contracts with a third party to independently verify results using the data and analysis materials submitted by the author (**https://ajps.org/ajps-verification-policy/**). Although the journal does not audit data for coding errors, this verification process is useful for eliminating computational and clerical errors, which seem to be widespread.

---

[6]Some have also warned about author bias and proposed methods of blind data analysis adopted in the hard sciences for countering it (MacCoun & Perlmutter 2015). These methods can be especially useful when replication is not feasible (MacCoun 2018) or when researchers analyze data in ways that go beyond preregistered plans (Dutilh et al. 2019).

Outside of the social sciences, some fields (e.g., cancer biology) are embracing post-publication peer review, which includes comments and critiques of published works, including errors found in those works (Horbach & Halffman 2018, p. 5; Vazire & Holcombe 2020). One service that enables post-publication review is PubPeer, which allows researchers to write comments that are attached to published articles through PubPeer's database and search mechanism (**https://pubpeer.com/**).

2.3. Assessing the Reform Efforts

An important component of the meta-research agenda is monitoring the effectiveness of reforms and initiatives (Hardwicke et al. 2020a, pp. 26–28). In this respect, some evidence suggests that the above responses are having their intended effect. Consider, for instance, self-reported adoption of replicability practices. A recent survey of political scientists, economics, and psychologists (the 3S Survey) found a substantial uptick in open materials, open data, and preregistration among researchers in those fields (Christensen et al. 2019b). This study also found that these upticks tend to align with the timing of new institutions and technologies aimed at promoting replicability.

Additional evidence of the effectiveness of recent reforms can be found in studies attempting to measure the effect of journal policies that require replicability (Hardwicke et al. 2018, Nuijten et al. 2017). In one study, researchers compared data availability in a judgment and decision-making journal that had adopted an open data policy to a comparator journal without such a policy (Nuijten et al. 2017, p. 14). They found a sharp increase in the replicability of articles published in the journal with the policy, but not in the journal that did not adopt it. That said, data made available in accordance with these policies are not always presented in a way that ensures others can easily make use of it (Hardwicke et al. 2018). In some cases, the policies are not enforced (Christensen & Miguel 2018, Hardwicke et al. 2018).[7] However, some evidence suggests that, to trigger compliance, these policies need not be framed as requirements. Awarding articles opt-in badges for practices like providing open data (e.g., an open data badge) is associated with journal-wide improvements (Kidwell et al. 2016), but seemingly only in fields in which the need for such reforms is already salient (Rowhani-Farid & Barnett 2018).

Despite substantial progress in both experimental and nonexperimental fields, especially in the last few years, much work remains to be done before the field of social science can claim to be benefiting fully from replicability policies (Christensen & Miguel 2018, Gelman 2017). Perhaps most concerning, journal editors resist publishing replication studies (Christensen & Miguel 2018). Most of the social sciences are, however, well ahead of law. In the next section, we turn to the state of progress in ELR.

3. THE STATE OF REPLICABILITY IN LAW

Calls for reform of replicability policies and practices in law go back to at least the early 2000s. Commentators seem to have valid cause for concern. Student-edited law journals regularly publish ELR, but only a few seek peer reviews, and journal editors generally do not have the requisite training to assess submission quality (Epstein & King 2002b). After summarizing long-standing and recently revived calls for reform, we provide a current snapshot of quality-control policies that top student- and faculty-edited law journals have adopted. We find that law lags behind other social science fields. Finally, we summarize relevant aspects of the first (to our knowledge) collective attempt to encourage and publish ELR replications. Our findings suggest that, although reform is underway, much more is needed.

---

[7]Even new methodology-focused papers do not always report their methods replicably (Obels et al. 2020).

3.1. Calls for Reform

Epstein & King (2002b, footnote 19) offer one of the earliest critiques of the quality of published ELR,[8] lamenting the "unmet need for a subfield of the law devoted to empirical methods...and methodological problems unique to legal scholarship." Their evaluation of all 231 articles with "empirical" in the title that were published in student-edited law journals between 1990 and 2000 uncovered rampant statistical errors—they claim that every study contained at least one.[9] They strongly endorse replicability as a necessary step to increase the quality of ELR, providing ample information about how to ensure complete disclosure of data and methods.[10] In the same vein, they recommend that law journal editors require authors to post all data and study materials to some public archive (p. 132). Others joined Epstein & King in early calls for replicability.[11] Mitchell (2004) lays out a detailed disclosure policy that law review student editors can cut and paste into their publication requirements.

More recently, several ELR scholars have revived the call to adopt replicability requirements and practices. Donohue (2015, pp. 318–21) emphasized the imperfections of the peer review process and called for open access to "the data and computer program" necessary to allow others to understand every step that goes into producing results.[12] He provided evidence of substantial methodological issues in published empirical studies of the effects of gun control laws on murder rates, the effects of the death penalty on deterrence, and other weighty issues. In 2016, four prominent intellectual property scholars published an open letter, signed by several other scholars, calling for, among other things, increased data and information disclosure in empirical intellectual property scholarship (Feldman et al. 2016). In the same year, Zeiler (2016) pointed to information disclosure as a key to ensuring the future relevance of ELR and called for existing academic societies to take active steps to push for and implement reforms. Irvine et al. (2018, pp. 346–48) proposed that experimental law and psychology researchers take steps to improve the replicability of their work.

A few legal scholars have raised questions about the value that the community places on replications and expressed concern over whether the culture of legal academia discourages the checking

---

[8]Unsurprisingly, the strongly critical article is controversial. See, e.g., Revesz (2002), who claims he made all information related to his criticized study available upon publication. In a reply, Epstein & King (2002b) push back on that claim, illustrating potential difficulties in verifying proper disclosure.

[9]They also report inferential errors in ELR published in faculty-edited journals.

[10]According to Epstein & King (2002a, p. 38),

> Good empirical work adheres to the replication standard: another researcher should be able to understand, evaluate, build on, and replicate the research without any additional information from the author. This rule does not actually require anyone to replicate the results of an article or book; it only requires that researchers provide information–in the article or book or in some other publicly available or accessible form—sufficient to replicate the results in principle.

Arlen & Talley (2008, p. xxxvi) propose the same for experimental legal research.

[11]In their ELR textbook, Epstein & Martin (2014) discuss the importance of replicable data sets and provide guidance on how to construct and document such data sets. In their textbook on empirical legal studies, Lawless et al. (2010, p. 40) emphasize the importance of replicable research. Spitzer (2003, p. 329) characterizes the push for transparency as "a superb idea." Mitchell (2004, p. 176) calls for law journal adoption of "a set of stringent disclosure requirements for reports of original empirical research, including disclosure of detailed information about methodology, data analysis, and the availability of raw data for replication and review." Nolasco et al. (2010) call for replicability in criminal justice research. Heise (2011, footnote 2) is amenable to replication by other scholars as a necessary feature of empirical legal scholarship.

[12]"Since such access is virtually never available to a referee, only a process of attempted replication could begin to reveal the many serious problems that can infect even major papers in top empirical journals" (Donohue 2015, p. 320).

of colleagues' work. Hubbard (2019, p. 1), for example, worries that successful replications might be "branded" by some, including journal editors, as "uninteresting." He advises researchers to "maximize their contribution to the field" by doing something more than merely attempting to replicate published results.[13] Although it is possible that this approach is optimal at least in the short term given current perceptions, we offer an alternative. Given what metascientists in the social sciences are unearthing about basic error rates, journal editors should consider standalone successful replications to be both surprising and novel and encourage such submissions. The production and publication of standalone replications is vital given the fallibility of the peer review process as a means of detecting, among other things, coding and computational errors (Nuijten et al. 2016). In fact, reform movement leaders in other fields have argued that "replication has more scientific value than original discovery" (Ioannidis 2018).

Hubbard (2019) also voices concern over "retaliation" by original authors whose work fails to replicate and the influence of such reactions on incentives to engage in the work of replication. We share Hubbard's concern and endorse the views of those who have pushed for reform despite feathers that might get ruffled.[14] The alternative eventually will lead to a complete lack of public trust in ELR research, especially if other social science fields continue to relentlessly press forward with quality-control reforms.

Despite the skepticism that Hubbard describes, calls for radical reform are growing in number and are now seemingly ubiquitous. We explore next whether the calls have been answered. We find that progress is severely lacking.

3.2. Law Journal Disclosure Requirements

We reviewed the transparency and openness guidelines of highly ranked faculty- and student-edited journals that publish legal research. For faculty-edited journals, we chose the top 30 by impact factor in the Web of Science's 2019 law database (omitting the student-edited journals in this list). For student-edited journals, we chose the top 30 in the 2019 Washington & Lee Law Journal Rankings as sorted by their primary metric (which considers citations on Westlaw). We searched the author guidelines for encouragement or requirement of open data, materials, or

---

[13]In his comments on a specific replication study, Hubbard (2019, p. 1) holds up the study as a model replication, noting that it "acknowledges the significance of the original study. . ., follow[s] in its footsteps even as it illuminates its methodological limitations and ultimately reaches very different empirical conclusions." He praises the replication study for its replicable methodological contributions that advance the state of knowledge. We note that the study Hubbard praises does not attempt to replicate the original results using the original data and methods. Instead, it proposes a new method for measuring a key variable, collects new data using that method, and employs them to produce results that lead to different conclusions. Although we agree that the new study makes an important and original contribution, we suggest that the study would have added something more useful to our knowledge base if it had performed a computational replication of the original results (Pashler & Harris 2012). Without a verification of the original results, we are left with a set of mixed results and a debate over which method for measuring the key variable is best. Had the authors started by attempting to replicate the original findings using the same data and same methods and found errors, they might have discovered that, once the errors were corrected, both measurement methods lead to the same results and that the measurement method is, in fact, irrelevant. Spellman & Kahneman (2018) argue that extensions of original research should routinely include replication. They recommend protocols to follow when extending experiments; the first step is replication. The benefits of routine replication apply in the case of nonexperimental research as well (Christensen & Miguel 2018).

[14]Dawn Chutkow suggests that senior scholars take the lead, either independently or jointly with junior scholars, given the "real concern of junior scholars, post-docs, and graduate students that [producing] replication studies that reveal meaningful problems with the original studies may not be career enhancing," and that faculty do more to utilize replication studies as pedagogical tools (email conversation on file with authors).

analytic code, and also for mentions of preregistration or accepting registered reports (i.e., some of the TOP Guidelines mentioned above). Full details of our search and the websites we reviewed are available in **Supplemental Tables 2 and 3**.

Beginning with the faculty-edited journals (**Table 2**), 14 of the 30 make no mention of any transparency guidelines. Several offer some encouragement (e.g., the journal encourages authors to post their data). Two journals (*International Environmental Agreements: Politics, Law and Economics* and the *Hague Journal on the Rule of Law*) require authors to include a statement

Table 2   Author guidelines at faculty-edited law journals[a]

Journal	Data	Materials	Code	Preregistration	Registered Reports	Impact factor
European Journal of Psychology Applied to Legal Context	0	0	0	0	0	4.91
International Data Privacy Law	0	0	0	0	0	4.16
Common Market Law Review	0	0	0	0	0	3.52
Regulation & Governance	0	0	0	0	0	3.38
Transnational Environmental Law	1	1	1	0	0	2.64
Annual Review of Law and Social Science	0	0	0	0	0	2.59
International Environmental Agreements: Politics, Law and Economics	2	2	2	0	0	2.51
American Journal of International Law	1	1	1	0	0	2.48
Journal of Law and the Biosciences	0	0	0	0	0	2.28
Antitrust Law Journal	0	0	0	0	0	2.26
Law and Human Behavior	0	0	0	0	Accepted	2.15
Feminist Legal Studies	0	0	0	0	0	2.10
Journal of International Economic Law	1	0	1	0	0	2.0
Computer Law & Security Review	3	1	1	0	0	1.85
Journal of Legal Analysis	0	0	0	0	0	1.73
Legal and Criminological Psychology	3	0	0	0	Accepted	1.67
Psychology, Public Policy, and Law	1	1	0	0	0	1.67
Journal of Environmental Law	1	0	1	0	0	1.63
Hague Journal on the Rule of Law	2	2	2	0	0	1.6
European Law Journal	1	1	1	0	0	1.59
International & Comparative Law Quarterly	1	1	1	0	0	1.57
European Constitutional Law Review	3	3	3	0	0	1.53
Modern Law Review	0	0	0	0	0	1.51
International Journal of Constitutional Law	0	0	0	0	0	1.51
World Trade Review	1	1	1	0	0	1.5
European Journal of International Law	0	0	0	0	0	1.48
Medical Law Review	1	0	1	0	0	1.46
International Journal of Transitional Justice	1	0	1	0	0	1.45
Law & Society Review	0	0	0	0	0	1.43
International Journal of Law and Psychiatry	1	0	1	0	0	1.34

[a]The top 30 faculty-edited journals law by impact factor in the Web of Science's Law category (2019). Author guidelines were examined for mentions of data, materials, and analytic code transparency, as well as preregistration and registered reports. Guidelines were coded for whether these factors were encouraged or required, or if a statement about them was required. 0 = no mention as a guideline; 1 = encourages sharing; 2 = requires an availability statement; 3 = requires sharing (subject to reasons why it is not possible). See **Supplemental Table 2** for more details and links to the journal websites that were reviewed.

Table 3 Author guidelines at student-edited law journals[a]

Journal	Data	Materials	Code	Preregistration	Registered Reports	W&L/impact factor
Yale Law Journal	3	0	3	0	0	1/6.59
Harvard Law Review	0	0	0	0	0	2/7.14
Stanford Law Review	3	0	0	0	0	3/3.37
Columbia Law Review	0	0	0	0	0	4/NR
University of Pennsylvania Law Review	0	0	0	0	0	5/4.62
Georgetown Law Journal	0	0	0	0	0	6/3.26
California Law Review	0	0	0	0	0	7/1.95
Notre Dame Law Review	0	0	0	0	0	8/1.23
University of Chicago Law Review	0	0	0	0	0	9/3.22
New York University Law Review	3	0	0	0	0	10/2.46
Iowa Law Review	0	0	0	0	0	11/1.0
UCLA Law Review	0	0	0	0	0	12/2.64
Minnesota Law Review	0	0	0	0	0	13/1.57
Texas Law Review	0	0	0	0	0	14/2.67
Michigan Law Review	0	0	0	0	0	15/3.53
Virginia Law Review	1	0	0	0	0	16/2.60
Duke Law Journal	0	0	0	0	0	17/3.08
William & Mary Law Review	0	0	0	0	0	18/NR
Vanderbilt Law Review	0	3	3	0	0	19/2.43
Fordham Law Review	0	0	0	0	0	20/1.02
Boston College Law Review	0	0	0	0	0	21/NR
Boston University Law Review	0	0	0	0	0	22/2.43
Cornell Law Review	0	0	0	0	0	23/1.66
Northwestern University Law Review	0	0	0	0	0	24/1.07
Cardozo Law Review	0	0	0	0	0	25/NR
Washington University Law Review	0	0	0	0	0	26/2.05
Southern California Law Review	0	0	0	0	0	27/1.60
U.C. Davis Law Review	0	0	0	0	0	28/NR
George Washington Law Review	0	0	0	0	0	29/1.24
Emory Law Journal	0	0	0	0	0	30/NR

[a]The top 30 student-edited journals according to the Washington & Lee (W&L) Law Journal Rankings (2019), based on its "Combined Score" metric. Impact factor from the Web of Science is also included. Guidelines were coded for whether these factors were encouraged or required, or if a statement about them was required. 0 = no mention as a guideline; 1 = encourages sharing; 2 = requires an availability statement; 3 = requires sharing (subject to reasons why it is not possible). Abbreviation: NR, not rated. For underlying data, see **Supplemental Table 3**.

Supplemental Material >

about the availability of their data, materials, and code, and just three (*Computer Law & Security Review*, *Legal and Criminological Psychology*, and *European Constitutional Law Review*) require some form of transparency or an explanation from the author justifying noncompliance. Two journals, both affiliated with psychology, accept registered reports (*Law and Human Behavior* and *Legal and Criminological Psychology*).

The picture at student-edited journals (**Table 3**) is less rosy. Of the 30 journals, 25 make no mention of any transparency guideline. Surprisingly, 4 of the 5 journals that do include such a mention are demanding about it. These journals (Yale, Stanford, New York University,

and Vanderbilt law reviews) require some aspect of replicability absent justification.[15] The fifth (*Virginia Law Review*) encourages open data. None mentions preregistration or registered reports. Unlike the faculty-edited journals, the student-edited journals that do have requirements appear clustered at the top of the rankings (they are more evenly distributed in the faculty-edited group).

Although not applicable to mere failures to present research in a replicable way, retraction policies can aid in ensuring that materially incorrect empirical results are permanently removed from the literature.[16] We searched the faculty- and student-edited journals for retraction policies on their websites.[17] All but three faculty-edited journals had retraction policies through their publisher (e.g., Oxford Academic's retraction policy applies to all of its journals). In contrast, no student-edited journal has a publicly available retraction policy.

3.3. Replication Projects in Empirical Legal Research

Although disclosure practices and policies are crucial, their value is substantially diminished if researchers fail to use the disclosed information to check for errors by attempting to replicate results. Suchman & Mertz (2010, p. 573) voiced concern more than a decade ago about the dearth of ELR replications using disclosed information.[18] Unfortunately, our findings suggest not much has changed over the past decade. Efforts similar to those in other fields have not made their way to law. For example, we are not aware of any large ELR-focused, multi-lab replication effort similar to those conducted in psychology and economics (**Table 1**).

One bright spot is a recently established annual conference focused on replication studies held by Claremont McKenna College's Program on Empirical Legal Studies (PELS) in 2018 and 2019.[19] To our knowledge, the conference is the first collective attempt at ELR replications. The *International Review of Law and Economics* has, so far, published 18 of the 20 studies presented at the 2018 and 2019 conferences.[20] Of these, we categorize 3 (Caspi & Stiglitz 2020, Kornhauser et al. 2020, Nyarko 2019) as completely new studies (i.e., neither replications nor robustness checks). Seven others test the robustness of published results by conducting alternative tests, employing either different data drawn from different populations or different methods to explore theories or conjectures that have found support in previously reported data. They do not attempt to replicate previously published results to identify coding, analytical, and statistical errors or to remedy sampling error or address low power.[21]

---

[15]In 2020, the *NYU Journal of Intellectual Property and Entertainment Law* instituted a practice of validating statistical findings before publication (Bass et al. 2020).

[16]It should be noted, however, that retracted papers seem to be cited frequently (Bolboacă et al. 2019).

[17]For **Tables 2** and **3** with information about retraction policies added, see **Supplemental Table 1**.

[18]Vazire & Holcombe (2020) propose a preliminary list of requirements for a system of correction that includes both transparency (information disclosure) and critical appraisal (checking each other's work).

[19]See Hoeppner (2019, p. 98), however, for a critique of the set of studies grounded in their general silence "about evaluating their replication results and, specifically, about how their results compare to the original results." He suggests that replication authors employ Bayesian methods to evaluate how their results impact inferences we can draw from the original study, the replication, and other previous findings.

[20]Our online materials include the 2018 agenda (**Supplemental Appendix 1**), the 2019 agenda (**Supplemental Appendix 2**), and a summary and categorization, using Clemens's (2015) taxonomy, of all 18 published studies (**Supplemental Appendix 3**).

[21]Five involve nonexperimental, observational data, and two attempt to test the robustness of results from experiments. Zwaan et al. (2018, p. 4) define an "alternative test" as a method used to "evaluate the robustness of a theoretical claim to alternative research designs." Using terminology proposed by Clemens (2015), these seven studies extend (using the same methods but different data drawn from a different population) or reanalyze (using the same data—or a different sample drawn from the same population—but different methods)

Most relevant for our purposes are the remaining 8 studies, which do attempt to replicate original results (Brady et al. 2019, Doleac et al. 2020, Hausladen et al. 2020, Ouellette & Tutt 2020, Prescott & Pyle 2019, Rao 2019, Sander 2019, Thorley & Mitts 2019). Of the 8, 7 involve nonexperimental, observational data drawn from the field, and 1 (Doleac et al. 2020) attempts to replicate results using data previously collected through 3 field experiments. In all, the 8 studies attempt to replicate results reported in 13 published studies.[22] Importantly, the authors of all 8 replication studies noted some level of difficulty in obtaining all materials required to perform a full replication, reflecting the downsides of the lack of formal journal policies and current customary practice. Perhaps more concerning, authors of only 2 of the 8 replication studies explicitly state that all of their data and analysis materials are publicly available.

The replication results are similar to what we have seen in other fields. The results are mixed, and some suggest major flaws that caution against applying, in any policy settings, the original study's purportedly supported claims until further study is possible. The single replication study focused on experiments successfully replicated the results reported by its three target studies using the original data sets but found that all three studies were underpowered—the original samples were insufficiently large to detect any existing effects (Doleac et al. 2020). Using methods to refine the results to account for low power, the authors reported that the conclusions from two of the three studies do not find support in the original data. Among the ten observational target studies, three were generally successfully replicated.[23] Two were found to have immaterial coding errors or immaterial differences between the original data set and the replicated data set;[24] three contained material coding errors;[25] two contained results that could not be fully replicated, suggesting computational or clerical errors;[26] and four had potential power issues.[27]

Outside of the PELS conference context, individual replication studies in law appear to be rare. Those that have been conducted, however, make important contributions.[28] We point here to just three examples, but there certainly are others. Williams et al. (2011) (reporting a failed replication attempt that prompted the original study's author to publish a correction) claimed to have cast significant doubt on the original study's claims. Irvine et al. (2018) (attempting replications along with extensions) provided only mixed results for the original findings. Hazelton et al. (2010) published a replicable replication upon request of an author under attack for publishing controversial and partially incorrect results related to the influence of campaign contributions on elected judges' decisions on cases. We note that unlike replications in other fields that we discuss above, almost none of the individual replication studies we found in law were preregistered and apparently replicable.

Although these findings demonstrate the importance of developing a culture of information sharing and result replication, drawing inferences from them about the general quality of ELR

---

previously published results and/or perform original studies (using different data and different methods). Although these are important aims, we do not count them as replications. Clemens uses replication to include verification and reproduction and robustness to include reanalysis and extension, so "studies that do not appear to be replications" is not strong enough.

[22] All eight replication studies report results from reanalyses, refinements, and/or extensions in addition to replications.

[23] Hausladen et al. (2020), Prescott & Pyle (2019), Thorley & Mitts (2019).

[24] Brady et al. (2019), Rao (2019) (Van Harten target study).

[25] Ouellette & Tutt (2020) (both target studies), Sander (2019) (Ayres and Brooks target study).

[26] Rao (2019) (Van Harten target study), Sander (2019) (Ayres and Brooks target study).

[27] Rao (2019) (both target studies), Sander (2019) (both target studies).

[28] A search to locate a complete set of replications is beyond the scope of our review. Note that some of the studies included here do not replicate original results but only reanalyze, extend, and/or perform new studies, which in our view is not ideal but is a move in the right direction.

is impossible. The target studies were not drawn randomly from the population of all published studies. Weaker studies might have been targeted, although some of the weakest studies likely cannot be replicated because necessary materials are unavailable. Moreover, authors may refrain from conducting replication research that is likely to confirm the original work, or submit work that has done so, because editors might prefer replications that call original results into question (Giner-Sorolla et al. 2018, Hubbard 2019).

4. CAN LAW FOLLOW SOCIAL SCIENCE?

Although legal researchers can benefit from and rely on the lessons and infrastructures being developed in other fields, law faces distinctive challenges in improving research replicability. This section briefly reviews those before exploring ways forward.

4.1. Challenges for Empirical Legal Research Reform Efforts

ELR is materially different from other research fields, and these differences pose significant barriers to reform. First, student editors stand as significant hurdles (Epstein & King 2002b). Law journals historically have operated independently, shielded from faculty influence. The potential for significant replicability reform adoption across a family of journals under, say, one publisher does not exist in law (Cent. Open Sci. 2020a). Also, student editors are unlikely to publish replication studies given their problematic preemption conventions (Moreno 2017, p. 423).

In addition, student editors do not face reputation effects that might influence faculty editors. Students serve on law journal editorial boards generally for one year. They face little risk of reputational hits when published authors refuse to disclose information used to produce empirical results and when results fail to replicate. Moreover, given that most law review student editors are not qualified to look for signs of errors in empirical research and that most law journals do not require reviews of work by qualified peers, the likelihood that obvious errors are published is quite high (Epstein and King 2002a).[29]

Second, the lack of faculty involvement on editorial boards is a barrier to adoption of policies designed to increase transparency. Student editors are likely to be unaware of methodological literatures related to transparency issues. Students are not on editorial boards long enough to generate any momentum for transparency innovations. Perhaps most problematic, students likely often select empirical articles not for their methodological soundness but rather because they wish to publish results that support their preferred descriptive or normative positions, or simply because they seem exciting.

Third, authorship norms may be standing in the way of improvements in replicability. This is because researchers trained primarily in legal methods may not have the knowledge to ensure their work is replicable. At the same time, norms in law seem to favor, at least compared to other fields, sole-authored articles or articles with few authors (Epstein & King 2002a). These restrictions may prevent legal researchers from involving methods specialists who can help improve the work's replicability (Holcombe 2019). More generally, a lack of diversity in the research process has been implicated as a source of low quality (Vazire & Holcombe 2020).

Fourth, authors that publish in law journals do not have robust incentives to disclose information necessary for replication. As we noted in Section 3, very few journal policies require such disclosure. Customs and norms in ELR do not seem to fill in the gaps by encouraging replicability.

---

[29]Of course, the peer review process is not a panacea. Donohue (2015) is one among many who have highlighted serious flaws in the academy's peer review process.

Without these incentives, authors may see replicability as simply added work that may open them up to scrutiny.

Finally, we suspect that reliance on eminence as a signifier of research quality (Vazire 2017) may be playing a role in perpetuating low-quality, irreplicable ELR. Some evidence suggests the absence of a clear link between an author's status and the quality of their work (Vazire 2017), and many practices and norms in ELR may overemphasize status in publication decisions. For instance, student editors do not evaluate articles blindly, so they are likely swayed by author status (see Harrison & Mashburn 2015, Yoon 2013).

4.2. Ideas for Reform

Methods for changing research cultures range from requiring replicable practices, to incentivizing them, to making them easier to implement (Nosek 2019). Beginning with more assertive measures, we have already discussed journal transparency and openness guidelines, and their general lack of adoption among legal journals. But, whereas law journals have been slow to adopt these guidelines (**Table 2** and **3**), some prominent journals have fairly strong data transparency guidelines.[30] These advances may help encourage others to follow suit. Similarly, increasing adoption outside of law may persuade editors to be more proactive in adjusting their guidelines. As to how those reforms may spread, there is some degree of centralization among North American law journals. For instance, there is a long-standing conference attended by law review editors (Closen & Jarvis 1992). And, in the past, these journals have worked together to attempt to collectively lower word counts in articles (Dorf 2005). Groups of empirical legal researchers (e.g., the Society for Empirical Legal Studies) may wish to circulate model guidelines, adapted from those used by other social sciences, for journals to consider and adopt. These guidelines should reflect best practices about ethical data sharing and how to handle articles that involve proprietary data sets (Klein et al. 2018a).

Turning to incentives, some empirical legal researchers may perceive the costs of conducting their work more replicably as high and the benefits as remote and amorphous. One remedy is to raise awareness of the existing rewards of replicable research. For instance, sharing data is associated with higher citation rates (Piowar & Vision 2013). Funders also recognize the importance of replicable practices by encouraging or requiring them (Natl. Acad. Sci. Eng. Med. 2018, p. 3). And, as we have discussed, the broader social scientific ecosystem increasingly views replicable methods as an important component of credible research. Law journals could also incentivize replicability by providing results-blind peer review under the registered reports model (Chambers 2019), offering opportunities for post-publication peer review (Horbach & Halffman 2018) and adopting badge systems (Kidwell et al. 2016, Rowhani-Farid & Barnett 2018).

Law schools and other institutions should also consider shifting incentives toward replicability. Law schools seem to acknowledge the value of empirical research when hiring (Zeiler 2016, p. 91); however, it is less clear that they consider the replicability of that work. Hiring committees can take these factors into account either through the applicant's published research or simply through asking about their practices. Such a policy has the further advantage of aligning a school's research interests with the growing movement away from research assessment based on traditional metrics

---

[30]In addition, at least one law journal invites replications for publication. The *Journal of Legal Studies* recently explicitly welcomed submissions that "reexamin[e] the results" of empirical studies previously published by the journal. The *Journal of Legal Studies* and a second University of Chicago law journal, the *Journal of Law and Economics*, also have recently adopted the *American Economic Review*'s data disclosure policy.

like impact factor and toward the quality and replicability of research (Moher et al. 2020).[31] Other school policies could also incentivize replicability, such as by asking faculty whether they plan to make their data and materials available when applying for internal funding.

Outside of law schools, the influential Washington & Lee Law Journal Rankings take into account court and academic citations to articles in the ranked journals. This methodology does not take into account replicability, despite the fact that for articles to properly be used in court, their methods and data should be fully examinable (Chin et al. 2019). To incentivize such practices, the rankers should consider including replicability as part of their system.

Turning to making replicable research easier and less daunting, groups like the Society for Empirical Legal Studies might consider using meetings to offer workshops and training on replicability methods. Organizations like the Framework for Open and Reproducible Research Training provide syllabi and other resources to assist in such initiatives. Empirical legal researchers might also find that, although earlier tools for sharing data and materials were initially designed for technical audiences (e.g., GitHub), more user-friendly websites have been developed over the past several years. Training and emerging best practice guidelines can also assist authors in navigating the ethical and legal issues that arise in making their work more accessible (Klein et al. 2018a).

Finally, ELR should join the meta-research movement that many cognate fields are engaged in (Hardwicke et al. 2020a). We hope our preliminary review of journal guidelines will help and can be revisited in the coming years to check for progress. And, as we saw in Section 3, many legal scholars conduct observational research, yet there do not seem to be any reporting guidelines tailored to this type of work, unlike in other fields that use observational data (e.g., medical researchers use the Strengthening the Reporting of Observational Studies in Epidemiology statement to encourage replicability in their field) (von Elm et al. 2008).[32] Development and testing of a consensus-based reporting statement for law and economics studies represents a straightforward way to improve ELR.

5. CONCLUSION

Replicability is foundational for scientific progress. Recent evidence demonstrating a serious lack of replicability of published social science research has spurred a robust and sustained movement to reform empirical research production procedures and publication requirements. The field of ELR faces significant obstacles to achieving necessary reforms. Thus, it is unsurprising that ELR lags behind other social scientific fields. This is troubling. ELR is designed to address questions of law and policy; its credibility is essential to these aims. Indeed, the field's credibility hinges on its willingness to implement mechanisms that detect and correct error.

Additional basic research is necessary to gain further insight into the extent of irreplicability of ELR. We do not yet have a sense of how widespread the problems are. Specifically, we need systematic study of the rates of inadequate disclosures; the rates of failure in replication attempts; whether journals that have adopted disclosure requirements strictly enforce them; whether consumers of ELR, including judges, legislators, regulators, and other legal actors, rely on reported findings despite potentially high error rates; and perceptions of the problems and potential solutions in the minds of authors and editors. So far, we have seen just the tip of the iceberg.

---

[31]The fact that US News and World Report just started ranking law schools based on citation counts, however, poses a significant challenge (Morse 2019).
[32]The Consolidated Health Economic Evaluation Reporting Standards seem to be followed in health research but not in economics (Christensen & Miguel 2018).

Most important, from our perspective, is the need for a culture shift in ELR. The field lacks coherence on fundamental issues, including what replication entails. It seems not to adequately value replications designed only to verify whether the author's data and methods actually support the reported results. LoPucki (2015, footnote 118) argues that requiring all legal empiricists to disclose information required for replicability would "make pervasive empiricism impossible." This might be true, but in our view, the potential cost of reduced quantity is well worth the benefit of improved quality. LoPucki (2015, abstract) also claims that one of the downsides of law schools hiring PhD-trained scholars is that they "impose the cultures of their disciplines on legal scholarship." In the case of replicability efforts, we believe ELR stands to benefit from the cultures of other social science fields.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

Aczel B, Szaszi B, Sarafoglu A, Kekecs Z, Kucharský Š, et al. 2020. A consensus-based transparency checklist. *Nat. Hum. Behav.* 4:4–6

Am. Econ. Assoc. 2020. *AEA RCT Registry*. **https://www.socialscienceregistry.org/**

Argyrou A. 2017. Making the case for case studies in empirical legal research. *Utrecht Law Rev*. 13:95–113

Arlen J, Talley EL. 2008. Introduction. In *Experimental Law and Economics*, ed. J Arlen, EL Talley, pp. xv–lxi. Cheltenham, UK: Edward Elgar

Bass ZJ, Ulrich AC, McLeod RB, Qiao K, Dynak J, et al. 2020. Editorial, the need for collective standards: validating raw data in legal empirical analysis. *NYU J. Intellect. Prop. Entertain. Law* 10:40–42

Bavli HJ. 2022. Credibility in empirical legal analysis. *Brooklyn Law Rev.* 87. In press

Beerdsen E. 2020. Litigation science after the knowledge crisis. *Cornell Law Rev*. 106:529–90

Bolboacă SD, Buhai D-V, Aluaş M, Bulboacă AE. 2019. Post retraction citations among manuscripts reporting a radiology-imaging diagnostic method. *PLOS ONE* 14:e0217918

Brady J, Evans MF, Wehrly EW. 2019. Reputational penalties for environmental violations: a pure and scientific replication study. *Int. Rev. Law Econ.* 57:60–72

Bruns SB, Ioannidis JPA. 2016. *p*-Curve and *p*-hacking in observational research. *PLOS ONE* 11:e0149144

Butler N, Delaney H, Spoelstra S. 2017. The gray zone: questionable research practices in the business school. *Acad. Manag. Learn. Educ.* 16:94–109

Camerer CF, Dreber A, Forsell E, Ho T-H, Huber J, Johannesson M, et al. 2016. Evaluating replicability of laboratory experiments in economics. *Science* 351:1433–36

Camerer CF, Dreber A, Holzmeister F, Ho T-H, Huber J, et al. 2018. Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nat. Hum. Behav.* 2:637–44

Caspi A, Stiglitz EH. 2020. Measuring discourse by algorithm. *Int. Rev. Law Econ.* 62:105863

Cent. Open Sci. 2020a. *APA joins as new signatory to TOP guidelines*. News Rel., Nov. 10. **https://www.cos.io/about/news/apa-joins-as-new-signatory-to-top-guidelines**

Cent. Open Sci. 2020b. *Transparency and openness guidelines*. **https://www.cos.io/our-services/top-guidelines**

Chambers C. 2019. What's next for Registered Reports? *Nature* 573:187–89

Chang AC, Li P. 2021. Is economics research replicable? Sixty published papers from thirteen journals say "often not". *Crit. Finance Rev.* 10:104. **http://dx.doi.org/10.1561/104.00000053**

Chin JM, Ribeiro G, Rairden A. 2019. Open forensic science. *J. Law Biosci.* 6:255–58

Christensen G, Freese J, Miguel EA. 2019a. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Oakland: Univ. Calif. Press

Christensen G, Miguel E. 2018. Transparency, reproducibility, and the credibility of economics research. *J. Econ. Lit.* 56:920–80

Christensen G, Wang Z, Paluck EL, Swanson N, Birke D, et al. 2019b. *Open science practices are on the rise: The State of Social Science (3S) Survey*. MetaArxiv. **https://doi.org/10.31222/osf.io/5rksu**

Clemens MA. 2015. The meaning of failed replications: a review and proposal. *J. Econ. Surv.* 31:326–42

Closen ML, Jarvis RM. 1992. The National Conference of Law Reviews Model Code of Ethics: final text and comments. *Marquette Law Rev*. 75:509–28

Collabra: Psychology. 2020. *Aims and scope*. **https://online.ucpress.edu/collabra/pages/About**

Dewald WG, Thursby JG, Anderson RG. 1986. Replication in empirical economics: the *Journal of Money, Credit and Banking Project*. *Am. Econ. Rev.* 76:587–603

Diamond SS, Mueller P. 2010. Empirical legal scholarship in law reviews. *Annu. Rev. Law Soc. Sci.* 6:581–99

Doleac JL, Temple C, Pritchard D, Roberts A. 2020. Which prisoner reentry programs work? Replicating and extending analyses of three RCTs. *Int. Rev. Law Econ.* 62:105902. **https://doi.org/10.1016/j.irle.2020.105902**

Donohue JJ. 2015. Empirical evaluation of law: the dream and the nightmare. *Am. Law Econ. Rev.* 17:313–60

Dorf MC. 2005. Thanks to a joint statement by top law journals, law review articles will get shorter, but will they get better? *FindLaw*, Feb. 28. **https://supreme.findlaw.com/legal-commentary/thanks-to-a-joint-statement-by-top-law-journals-law-review-articles-will-get-shorter-but-will-they-get-better.html**

Dutilh G, Sarafoglou A, Wagenmakers EJ. 2019. Flexible yet fair: blinding analyses in experimental psychology. *Synthese*. **https://doi.org/10.1007/s11229-019-02456-7**

Ebersole CR, Atherton OE, Belanger AL, Skulborstad HM, Allen JM, et al. 2016. Many Labs 3: evaluating participant pool quality across the academic semester via replication. *J. Exp. Soc. Psychol.* 67:68–82

Epstein L, King G. 2002a. The rules of inference. *Univ. Chicago Law Rev.* 69(1):1–133

Epstein L, King G. 2002b. A reply. *Univ. Chicago Law Rev.* 69:191–210

Epstein L, Martin AD. 2014. *An Introduction to Empirical Legal Research*. Oxford, UK: Oxford Univ. Press. 1st ed.

Feldman R, Lemley M, Masur J, Rai AK. 2016. Open letter on ethical norms in intellectual property scholarship. *Harvard J. Law Technol.* 29:1–14

Franco A, Malhotra N, Simonovits G. 2015. Underreporting in political science survey experiments: comparing questionnaires to published results. *Political Anal.* 23:306–12

Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. 2018. Questionable research practices in ecology and evolution. *PLOS ONE* 13:e0200303

Freese J, Peterson D. 2017. Replication in social science. *Annu. Rev. Sociol.* 43:147–65

Gelman A. 2017. Ethics and statistics: Honesty and transparency are not enough. *Chance* 30:37–39

Gentzkow M, Shapiro JM. 2014. *Code and data for the social sciences: a practitioner's guide*. Mimeo, Univ. Chicago. **http://web.stanford.edu/~gentzkow/research/CodeAndData.pdf**

Giner-Sorolla R, Amodio DM, van Kleef GA. 2018. Three strong moves to improve research and replications alike. *Behav. Brain Sci.* 41:e130

Glandon P. 2010. *Report on the American Economic Review Data Availability Compliance Project*. Publ., Kenyon Coll., Gambier, OH. **http://digital.kenyon.edu/cgi/viewcontent.cgi?article=1011&context=economics_publications**

Han S, Olonisakin TF, Pribis JP, Zupetic J, Yoon JH, et al. 2017. A checklist is associated with increased quality of reporting preclinical biomedical research: a systematic review. *PLOS ONE* 12:e0183591

Hardwicke TE, Mathur MB, MacDonald K, Nilsonne G, Banks GC, et al. 2018. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*. *R. Soc. Open Sci.* 5:180448

Hardwicke TE, Serghiou S, Janiaud P, Danchev V, Crüwell S, et al. 2020a. Calibrating the scientific ecosystem through meta-research. *Annu. Rev. Stat. Appl.* 7:11–37

Hardwicke TE, Wallach JD, Kidwell MC, Bendixen T, Crüwell S, Ioannidis JPA. 2020b. An empirical assessment of transparency and reproducibility-related research practices in the social sciences. *R. Soc. Open Sci.* 7:190806

Harrison JL, Mashburn AR. 2015. Citations, justifications, and the troubled state of legal scholarship: an empirical study. *Texas A&M Law Rev.* 3:45–90

Hausladen CI, Schubert MH, Ash E. 2020. Text classification of ideological direction in judicial opinions. *Int. Rev. Law Econ.* 62:105903

Hazelton MW, Hinkle RK, Martin AD. 2010. On replication and the study of the Louisiana Supreme Court. *Glob. Jurist* 10:85–91

Heise M. 2011. An empirical analysis of empirical legal scholarship production, 1990–2009. *Univ. Ill. Law Rev.* 2011:1739–52

Hoeppner S. 2019. A note on replication analysis. *Int. Rev. Law Econ.* 59:98–102

Holcombe A. 2019. Farewell authors, hello contributors. *Nature* 571:147

Horbach SPJM, Halffman W. 2018. The changing forms and expectations of peer review. *Res. Integr. Peer Rev.* 3:8

Howe PDL, Perfors A. 2018. An argument for how (and why) to incentivise replication. *Behav. Brain Sci.* 41:e135

Hubbard WHJ. 2019. A replication study worth replicating: a comment on Salmanowitz and Spamann. *Int. Rev. Law Econ.* 58:1–2

Hubbard WHJ, Hyman D. 2019. What's in a name? A taxonomy of replication. *Int. Rev. Law Econ.* 60:1–3

Ioannidis JPA. 2005. Why most published research findings are false. *PLOS Med.* 2:e124

Ioannidis JPA. 2018. Why replication has more scientific value than original discovery. *Behav. Brain Sci.* 41:e137

Irvine K, Hoffman D, Wilkinson-Ryan T. 2018. Law and psychology grows up, goes online, and replicates. *J. Empir. Leg. Stud.* 15(2):320–55

John LK, Loewenstein G, Prelec D. 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* 23:524–32

Kidwell MC, Lazarević LB, Baranski E. 2016. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLOS Biol.* 14:e1002456

Klein O, Hardwicke TE, Aust F. 2018a. A practical guide for transparency in psychological science. *Collabra Psychol.* 4:20

Klein RA, Ratliff KA, Vianello M, Adams RB Jr., Bahník S. 2014. Investigating variation in replicability a "many labs" replication project. *Soc. Psychol.* 45:142–52

Klein RA, Vianello M, Hasselman F, Adams BG, Adams RB, et al. 2018b. Many Labs 2: investigating variation in replicability across samples and settings. *Adv. Meth. Pract. Psychol. Sci.* 1:443–90

Koenker R, Zeileis A. 2009. On reproducible econometric research. *J. App. Econ.* 24:833–47

Kornhauser L, Lu Y, Tontrup S. 2020. Testing a fine is a price in the lab. *Int. Rev. Law Econ.* 63:105931

Kvarven A, Strømland E, Johannesson M. 2020. Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nat. Hum. Behav.* 4:423–34

Lawless RM, Robbennolt JK, Ulen T. 2010. *Empirical Methods in Law*. New York: Aspen Publ.

Learner EE. 1983. Let's take the con out of econometrics. *Am. Econ. Rev.* 73:31–43

LoPucki LM. 2015. Disciplining legal scholarship. *Tulane Law Rev.* 90:1–34

MacCoun R. 2018. Enhancing research credibility when replication is not feasible. *Behav. Brain Sci.* 41:e142

MacCoun R. 2021. *p*-hacking: a strategic analysis. In *Research Integrity in the Behavioral Sciences*, ed. L Jussim, ST Stevens, JA Krosnick. New York: Oxford Univ. Press. In press

MacCoun R, Perlmutter S. 2015. Hide results to seek truth. *Nature* 526:187–89

Makel MC, Hodges J, Cook BG, Plucker JA. 2021. Both questionable and open research practices are prevalent in education research. *Educ. Res.* 1. **https://doi.org/10.3102/0013189X211001356**

McCullough BD. 2009. Open access economics journals and the market for reproducible economic research. *Econ. Anal. Policy* 39:117–26

Mitchell G. 2004. Empirical legal scholarship as scientific dialogue. *N.C. Law Rev.* 83:167–204

Miyakawa T. 2020. No raw data, no science: another possible source of the reproducibility crisis. *Mol. Brain* 13:24

Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, et al. 2020. The Hong Kong Principles for assessing researchers: fostering research integrity. *PLOS Biol*. 18:e3000737

Moher D, Liberati A, Tetzlaff J, Althman DG, PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLOS Med*. 6:e1000097

Moreno JA. 2017. 99 problems and the bitchin' is one: a pragmatist's guide to student-edited law reviews. *Touro Law Rev*. 33:407–30

Morse R. 2019. *U.S. News* considers evaluating law school scholarly impact. *U.S. News Blog*, Feb. 13. **https://www.usnews.com/education/blogs/college-rankings-blog/articles/2019-02-13/us-news-considers-evaluating-law-school-scholarly-impact**

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, et al. 2017. A manifesto for reproducible science. *Nat. Hum. Behav*. 1:0021

Natl. Acad. Sci. Eng. Med. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: Natl. Acad. Press

Natl. Acad. Sci. Eng. Med. 2019. *Reproducibility and Replicability in Science*. Washington, DC: Natl. Acad. Press

Necker S. 2014. Scientific misbehavior in economics. *Res. Policy* 43:1747–59

Nolasco CARI, Vaughn MS, del Carmen RV. 2010. Toward a new methodology for legal research in criminal justice. *J. Crim. Justice Educ*. 21:1–23

Nosek BA. 2019. Strategy for culture change. *Center for Open Science Blog*, June 11. **https://www.cos.io/blog/strategy-for-culture-change**

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, et al. 2015. Promoting an open research culture. *Science* 348:1422–25

Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. 2018. The preregistration revolution. *PNAS* 115:2600–6

Nuijten MB, Borghuis J, Veldkamp CLS, Dominguez Alvarez L, van Assem MALM, et al. 2017. Journal data sharing policies and statistical reporting inconsistencies in psychology. *Collabra Psychol*. 3:31

Nuijten MB, Hartgerink CHJ, van Assen MALM, Epskamp S, Wicherts J. 2016. The prevalence of statistical reporting errors in psychology (1985–2013). *Behav. Res*. 48:1205–26

Nyarko J. 2019. We'll see you in...court! The lack of arbitration clauses in international commercial contracts. *Int. Rev. Law Econ*. 58:6–24

Obels P, Lakens D, Coles NA, Gottfried J, Green SA. 2020. Analysis of open data and computational reproducibility in Registered Reports in psychology. *Adv. Meth. Pract. Psychol. Sci*. 3:229–37

Open Sci. Collab. 2015. Estimating the reproducibility of psychological science. *Science* 349:aac4716

Ouellette LL, Tutt A. 2020. How do patent incentives affect university researchers? *Int. Rev. Law Econ*. 61:105883

Pashler H, Harris CR. 2012. Is the replicability crisis overblown? Three arguments examined. *Perspect. Psychol. Sci*. 7:531–36

Piowar HA, Vision TJ. 2013. Data reuse and the open data citation advantage. *PeerJ* 1:e175

Prescott JJ, Pyle B. 2019. Identifying the impact of labor market opportunities on criminal behavior. *Int. Rev. Law Econ*. 59:65–81

Racine JS. 2019. *Reproducible Econometrics Using R*. Oxford, UK: Oxford Univ. Press

Rao W. 2019. Development status and decision-making in investment treaty arbitration. *Int. Rev. Law Econ*. 59:1–12

Revesz RL. 2002. A defense of empirical legal scholarship. *Univ. Chicago Law Rev*. 69:169–90

Rowhani-Farid A, Barnett AG. 2018. Badges for sharing data and code at *Biostatistics*: an observational study. *F1000Research* 7:90

Sander R. 2019. Replication of mismatch research: Ayres, Brooks and Ho. *Int. Rev. Law Econ*. 58:75–88

Simmons JP, Nelson LD, Simonsohn U. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci*. 22:1359–66

Simons DJ. 2018. Introducing advances in methods and practices in psychological science. *Adv. Methods Pract. Psychol. Sci*. 1:3–6

Spellman BA, Kahneman D. 2018. What the replication reformation wrought. *Behav. Brain Sci*. 41:e149

Spitzer M. 2003. Evaluating valuing empiricism (at law schools). *J. Leg. Educ.* 53:328–31

Suchman MC, Mertz E. 2010. Toward a new legal empiricism: empirical legal studies and new legal realism. *Annu. Rev. Law Soc. Sci.* 6:555–79

Thorley D, Mitts J. 2019. Trial by Skype: a causality-oriented replication exploring the use of remote video adjudication in immigration removal proceedings. *Int. Rev. Law Econ.* 59:82–97

Vanpaemel W, Vermorgen M, Deriemaecker L, Storms G. 2015. Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra* 1:3

Vazire S. 2017. Our obsession with eminence warps research. *Nature* 547:7

Vazire S. 2018. Implications of the credibility revolution for productivity, creativity, and progress. *Perspect. Psychol. Sci.* 13:411–17

Vazire S. 2019. A toast to the error detectors. *Nature* 577:9

Vazire S, Holcombe AO. 2020. Where are the self-correcting mechanisms in science? PsyArXiv. kgqzt. **https://doi.org/10.31234/osf.io/kgqzt**

Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, et al. 2014. The availability of research data declines rapidly with article age. *Curr. Biol.* 24:94–97

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. 2008. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J. Clin. Epidemiol.* 61:344–49

Williams D, Sander R, Luppino M, Bolus R. 2011. Revisiting law school mismatch: a comment on Barnes (2007, 2011). *Northwest. Law Rev.* 105:813–28

Yoon A. 2013. Editorial bias in legal academia. *J. Leg. Anal.* 5:309–38

Zeiler K. 2016. The future of empirical legal scholarship: Where might we go from here? *J. Leg. Educ.* 66:78–99

Zwaan RA, Etz A, Lucas RE, Donnellan MB. 2018. Making replication mainstream. *Behav. Brain Sci.* 41:e120