

REPLICATION IN SOCIAL SCIENCE

Jeremy Freese

Department of Sociology, Stanford University

jfreese@stanford.edu

David Peterson

Department of Sociology, Northwestern University

dpeterson@u.northwestern.edu

Keywords: Replication, transparency, sociology of science, sociology of knowledge, quantitative methodology, qualitative methodology

Direct correspondence to: Jeremy Freese, Department of Sociology, Stanford University, 450 Serra Mall, Stanford, CA 94305-2047, *jfreese@stanford.edu*

Article table of contents

Abstract

Introduction

Sociology of Science on Replication

Forms of Replication in Quantitative Social Science

Verifiability

Robustness

Repeatability

Generalization

Replication in Qualitative Social Science

Conclusion

The Routine Ambiguities

Replication Policy in Sociology

Abstract

Across the medical and social sciences, new discussions about replication have been transforming research practices. Sociologists, however, have been largely absent from these discussions. The goals of this review are to introduce sociologists to these developments, synthesize insights from science studies about replication in general, and detail the specific issues regarding replication that occur in sociology.

The first half of the article argues that a sociologically sophisticated understanding of replication must take into account both the ways that replication rules and conventions evolved within an epistemic culture and how those cultures are shaped by specific research challenges. The second half outlines the four main dimensions of replicability in quantitative sociology- verifiability, robustness, repeatability, and generalizability- and discusses the specific ambiguities of interpretation that can arise in each. We conclude by advocating some commonsense changes to promote replication while acknowledging the epistemic diversity of our field.

Science has been said to be presently in the throes of a “crisis of replication.” NIH director Francis Collins described recent concerns about replicability as “a cloud over biomedical research” (Hughes 2014). In 2014, *Science* and *Nature*, along with 30 other journals, issued a nearly unprecedented joint editorial to announce new reproducibility initiatives that seek to “raise the standards... as part of the quality control that justifies public trust in science” (Nature 2014). In behavioral science, psychology has had a series of dramatic episodes surrounding replication (including one with the catchy hashtag “#repligate”), provoking a movement whose recent influence in social psychology has been described as “staggering” (Finkel et al. 2015). Economics and political science have both advanced their own initiatives to address “transparency,” “reproducibility,” and “replication.”

Amidst these transdisciplinary developments, sociology’s own institutions and research practices remain relatively unaffected. A passing glance at sociology might even suggest that replication has become a vanishing art: sociology’s two flagship journals have published 27 articles over their history with “replication” in the title, but only one in the last 25 years. What makes this especially ironic is that sociology not only provided some of the most compelling early statements about replication in social science (e.g., Rose 1953; Mack 1951) but, through the sociology of science, the discipline has arguably contributed the most of any social science to understanding how replication actually works in practice.

In recognition of both traditions, we attempt here to advance a sociologically self-aware discussion of replication in social research: that is, a review of issues regarding replication in social science that is informed by and consistent with the contributions science studies has made regarding scientific cultures and replication. We begin by synthesizing what we think are the key insights that sociology of science has to offer discussions of replication. We then use these ideas to organize and consider some issues that have emerged in discussions about replication in quantitative social science. We spend the bulk of the review on quantitative social science because qualitative inquiry poses very different issues about which replication may not even be the best term. We do discuss these issues in relation to qualitative social science later in the paper.

SOCIOLOGY OF SCIENCE ON REPLICATION

Replication played a central role in the theories of science popular in the mid-twentieth century. Philosophically, replication was considered essential to science because theories could be “falsified only if we discover a *reproducible* effect which refutes the theory” (Popper 1959:66, emphasis ours). Sociologically, that one’s work could be replicated provided a social control mechanism that discouraged shoddy or fraudulent work and was articulated as part of the norm of “organized skepticism” taken as key to the successful functioning of science (Merton 1973; Zuckerman 1977).

However, empirical studies of replication attempts and controversies in the natural sciences presented a more complex image. For one, many indisputably successful sciences rarely perform exact replications (Hacking 1983). Replication often involves craft knowledge that is not in published reports (Collins 1974; 2001; Mulkay 1984). Even for experts, original studies often do not provide enough information for replications to occur (Travis 1981). Instead of a neutral verification process, replication is sometimes used strategically to undermine adversaries (Pinch 1979).

This work makes clear that replication in practice does not operate the way an abstract “standard view” of science suggests. But, then, how should replication be understood? For us, a sociologically consistent view of replication in social science emphasizes four things. First, replication contains a series of unavoidable interpretive ambiguities. Second, these ambiguities are partially rooted in a tension between the epistemic value of similarity in replication and the value of difference. Third, scientific communities develop rules and conventions around replication that are meant to alleviate these ambiguities but can themselves become objects of epistemic debate. Finally, because these rules and conventions develop in response to both internal and external pressures, it is important to recognize the particular challenges that each epistemic culture faces rather than accept a universal theory of scientific replication.

Ambiguity and interpretation in replication

Consider the case in which a scientist sets out to determine if a published experiment can be replicated. If the replication fails, there might be a problem with the original result, but there might be something wrong with the replication. Or perhaps the failure is caused by some subtle difference that, if understood, would represent a scientific discovery in its own right.

Given this ambiguity, one might conclude that we can never know whether a replication study has been properly done unless it “works.” But this would imply that replications cannot be properly considered to be testing anything, if they are only interpreted as evidence when successful. So we need some way of judging the quality of an experiment that is independent of its outcome.

This circle is known as the “experimenter’s regress,” described by Harry Collins, whose work over more than four decades represents science studies’ most significant contribution to understanding replication (e.g., Collins 1974; 1985; 1991; 2001; 2016). To Collins, judgments about experiments are necessarily interpretive matters, and the prevailing interpretations of results in a field is the outcome of social interactions among scientists.

The conclusion fits the broad position of constructivist science studies that “there can never be strictly logical grounds for forsaking a theory” (Yearley 2005: 30). If logic alone is never sufficient, then it falls to sociologists and historians to help understand how scientific controversies are eventually resolved. If replication is always potentially subject to doubt, it falls to sociologists and historians to “show how communities come to discriminate between reasonable and unreasonable doubts” (Panofsky 2014: 160). Scholars in science and technology have continued to explore issues of replication by extending Collins’ framework into new areas (e.g., Kennefick 2000) or highlighting the new challenges for replication brought about by current technological or institutional developments in scientific practice (e.g., Edwards et al. 2011; Leahey 2008).

The constructivist position on replication has been the target of criticism by scholars who emphasize that replication—even if infrequently done, technically challenging, and ambiguous—plays an irreducible role in shaping scientific arguments (Cole 1992; Kitcher 1995). While Collins may dispute the independent power that replications hold, he admits that the ideal of replication is a lodestar, not just for natural science but even in social sciences where replication holds additional challenges. Thus, he argues that “The job of the social sciences is not to show that replication is futile or impossible, but to show how to pursue replicability in the face of its recalcitrance” (Collins 2016:78). For Collins, even as replication involves interpretive ambiguities and social negotiations, it remains “the only criterion [we have] of what is to count as a natural regularity (or social regularity)” (quoted in Ashmore 1989:138).

Similarity versus difference in replication

Central to Collins’ contribution is his demonstration that replications are characterized by an unavoidable tension between the perceived epistemic merits of maintaining similarity to the original study and the perceived merits of difference. Everyone recognizes that one cannot conduct exactly the same experiment on exactly the same subjects at the same time, but differences are sometimes understood as a *limitation* to be minimized and other times as a *virtue*.

In Table 1, we illustrate how similarity/difference in replication design and success/failure in replication outcome affect both narrow and broad interpretations of a finding. Here, a “narrow” interpretation of a finding sticks closely to the literal results and conditions of the study, while a “broad” interpretation speaks in terms of whatever theory or larger generalization the finding may be taken to support. When a replication attempt is extremely similar to the target study, it adds little additional information beyond evidence that the result of the first experiment was not simply a fluke (Collins 1984:34). It will not, for instance, aid in adjudicating between competing hypotheses. In terms of Table 1, success strengthens the original results in a narrow sense but, beyond this, does not add anything to our confidence of a broader interpretation of the findings.

[TABLE 1 ABOUT HERE]

The more ambiguous situation occurs when a replication designed to mimic the original study *fails*. As previously mentioned, this may cast doubt on both the original study's literal findings and the broader implications those findings were taken to imply. However, as we will outline, authors of an original study will often discount such failures by arguing that critical changes were introduced that vitiate its interpretation.

As a replication introduces greater contrast from the original experiment—for example, if it is performed in a new location or using new methods—more information may be gained. When a replication that differs in key respects from a target experiment fails to replicate, it is easy for authors to attribute failure to those differences because the replication knowingly altered the original design.

More ambiguity arises when such a replication *succeeds*. On one hand, the broader interpretation is strengthened, which also increases confidence that the original study was correct. Yet, as already noted, as a replication becomes more dissimilar, it becomes increasingly difficult to interpret. Thus, critics can argue that such replications are too different and no longer speak to the original finding.

Across science, debates over replication arise along the similarity/difference, success/failure axes. This has proved a useful tool for social scientists who study scientific conflict. Yet, this general principle says little about the specific ways that scientific fields manage such debates. Additionally, because most science studies work on the topic has concentrated on experimental research, it may say little about the particular issues that arise in social sciences in which observational studies predominate.

Epistemic cultures and replication

Among the fundamental shifts that the “second wave” (Collins & Evans 2002) of science studies brought was to stop viewing “science” as a unitary activity. As Shapin (2010:5) explains, scholars have

come to believe that “science is not one, indivisible, and unified, but ... many, diverse, and disunified.” Even a seemingly stable notion like “objectivity” has evolved different historical forms (Daston 1992; Daston & Galison 2010; Porter 1995). Science studies has highlighted how scientific communities are characterized by methods, theories, and practices that developed in path-dependent ways, yielding unique “epistemic cultures” (Galison & Stump 1996; Hacking 1996; Knorr Cetina 1999).

Thus, the role of replication within a field ought to be understood as the outcome of a process of cultural development which is influenced by both internal dynamics and external pressures rather than a universal feature of an idealized scientific method. Epistemic cultures may include *replication rules*, which are explicitly stated policies regarding replication expressed by professional associations, journals, and funding agencies. These include journal rules about what information about a study must be reported and rules surrounding sharing data and materials.

Around such rules, however, *replication conventions* involve all non-codified knowledge about replication. Examples include shared understandings of what it is appropriate to ask another researcher for, what studies are worthy of replicating, when replication becomes a considered an attack, and so on. Within a field, of course, researchers may have different ideas about what these conventions are, and that such differences exist may only become recognized in episodes where some perceive a breach.

The success of scientific communities in achieving their ostensible epistemic goals is up for debate by both those within a field and outsiders. When the replication rules and conventions of a field are deemed problematic, activists may seek to change them. Recent years have seen so much activism around replication as to be regularly characterized as a “movement” (Finkel et al 2015). This has involved efforts to change existing rules, to change their colleagues’ view of conventions, and, especially, to codify matters that had been left to convention as explicit rules.

Integrative replication

In fields characterized by rapid technological development, replication is often piecemeal, implicit and informal. As new techniques and technologies are developed by other labs, there are quick attempts to integrate them into the practice in order to stay at the cutting edge of a receding horizon of experimental possibilities, a process recently referred to as “bench-building” (Peterson 2015). For instance, if a molecular biologist publishes a finding based upon a new strain of mouse she developed, other molecular biologists may ask to use the mouse, not to explicitly replicate the experiment in order to evaluate the factual statements, but in order to “play with” and, possibly, integrate the technology into their practice. In this case, replication occurs as a byproduct of bench-building rather than an explicit goal in itself.

This form of “integrative replication” represents the major form that replication takes in laboratory science. This is why Hacking (1983) calls the focus on repeatability a “philosophical pseudo-problem” because “roughly speaking, no one ever repeats an experiment. Typically, serious repetitions of an experiment are attempts to do the same thing better- to produce a more stable, less noisy version of the phenomenon” (231). For Hacking, most replication occurs as a natural product of developing the technological capacity of the lab rather than as an explicit test for truth. Hacking famously distinguished “intervening” from “representing,” arguing that the philosophers have been too concerned with whether scientific facts properly represented nature while ignoring the importance of actually intervening in nature (an argument furthered in the sociology of science by Pickering 1995).

In fields that rarely intervene—the situation of the vast majority of social science—replication is instead necessarily concerned with the accuracy of representations. The difference to us seems not to reflect any immaturity of social science but inherent differences in subject matter: social science is not centered on the development and deployment of (literal) machinery to (literally) manipulate (literal) material objects. We emphasize this here because so much discussion of what social scientists should be doing proceeds from the premise that natural science provides the best role model. For replication, we think the most replication work in the natural scientists is simply incommensurate with what social

scientists are trying to do, resulting in very different replication rules and conventions. Consequently, normative questions about replication for social scientists need to be understood on their own terms.

FORMS OF REPLICATION IN QUANTITATIVE SOCIAL SCIENCE

As with the natural sciences, “replication” in quantitative social science has been used to characterize a variety of different activities. Clemens (2015) impressively catalogues a crazy quilt of overlapping and not-infrequently incommensurate terminological distinctions that past authors have presented regarding replication. Our discussion here is organized by whether the replication attempt is based upon the original data or gathers new data and the tension between similarity and difference described above.

We end up with four analytically distinct categories (Table 2).

[TABLE 2 ABOUT HERE]

(1) *Verifiability* involves taking the results of an original study as the object of inquiry and typically asks limited questions regarding whether the same results are obtained by doing the same analyses on the same data. (2) Tests of *robustness* conduct a reanalysis on the original data using alternative specifications to see if the target finding is merely the result of analytic decisions. (3) Tests of *repeatability* involve collecting new data to determine whether key result of a study can be observed by using the original procedures to collect new data. Finally, (4) for inquiries into *generalizability*, the original study provides a premise for other work that trying to evaluate whether similar findings may be consistently observed across different methods or settings.

Verifiability

We use *verifiability* to describe activities directed to interrogating whether the findings presented in a study follow properly from the study's data. The most straightforward version is someone determining whether they can reproduce results using what are understood to be the same data and same code. For some, this might seem so pedestrian as to not really count as replication (Collins 1991); to others, it is "pure replication" (Brown et al. 2014). Either way, the recent experience of *American Journal of Political Science (AJPS)* in actually using a third-party to do this for articles it publishes indicates that it is not to be taken for granted. The first fifteen articles all required at least some additional back-and-forth after data and code supposedly sufficient to verify results were submitted (Political Science Replication 2015).

Even when results can be reproduced, there is the question of whether data or code contain errors. A graduate student discovered that findings by Reinhart and Rogoff (2010) that had been used to support austerity policies were artifacts of a spreadsheet error (Herndon et al. 2014). Even if all the analyses are correct, there is the possibility of problems in the raw data. In quantitative social science, data are usually not collected by a study's authors, and a complicated chain of actors may separate the origins of data from its analysis. For instance, skepticism of findings by McPherson et al. (2006) about increasing social isolation in the United States led to the discovery of 41 people which the survey center providing the data had incorrectly coded as having no confidants (McPherson et al 2008).

Independent verification of results is only possible if data are available to others. Researchers who are able to keep anyone else from verifying results are insulated against any definitive demonstration of error. Efforts to verify results after publication may be the most interpersonally fraught type of replication because they are hard not to be interpreted as signaling distrust, with the possibility of the researcher's honesty or competence called into dispute. As described in Tischler (2007), Peterson (1996a, 1996b) had been thwarted for several years in his effort to verify Weitzman's widely reported finding that women's standard of living declined over 70% in the year following divorce. Peterson was allowed access only after intervention from the National Science Foundation, and found that the finding could not be reproduced (his own estimate was 27%).

The convention in sociology has largely been to treat the availability of data for verification as an ethical matter to be handled by individual requests. Systematic inquiries in other disciplines have found widespread failures of investigators to respond to post-publication data requests (McCullough et al. 2006; Wicherts et al. 2006; Parish et al. 2015). In sociology, Young and Horvath (2015) reported more than half of authors have refused to provide data or code to graduate students trying to verify 53 findings as a class exercise. In economics, a systematic inquiry revealing poor response to data requests led the *American Economic Review* to impose data sharing guidelines at the time of publication and monitor authors' compliance with them (Bernanke 2004). That is, key economists regarded replication conventions about data availability to be inadequate, and the discipline has since moved toward more formal replication rules.

Replication rules by journals stipulating transparency regarding data availability at the time of publication have spread over the past decade in social sciences other than sociology. Authors are usually urged to deposit data in independent repositories, developed and maintained by archiving professionals, rather than leaving long-term data availability contingent on authors maintaining their personal websites. Major repositories include the Open Science Framework, OpenICPSR, and Harvard's Dataverse initiative.

Although activists sometimes treat "openness" as an unambiguous good, requiring social scientists to make data available as a condition of publication raises various concerns about confidentiality, proprietary sources, and compromising future intellectual property rights of investigators. The latter is not unique to social science: a *New England Journal of Medicine* editorial noted concerns that clinical research could be taken over by "research parasites" whose reliance on other's data was "possibly stealing from the research productivity planned by the data gatherers" (Longo and Drazen 2016).

Numerous ideas to address these concerns have been advanced, including redactions, embargos, and improved data citation practices to give those who collect data greater credit when they share it.

More fundamentally, however, it seems almost certain that any feasible mandate for sharing data upon publication in social science, however, must include possibilities for exemption. In the *American Economic Review*, one-third of its articles in 2014 were exempted from its data-sharing requirements (Goldberg 2015). When authors explicitly articulate why data cannot be made available, the justifications may be evaluated by reviewers or approved by editors. Major political science journals promote providing information about data availability in the first footnote (starting with Meier 1995). The larger point is that good reasons may be given for why data cannot be made publicly available, but it is hard to argue against expecting researchers to be transparent about what the conditions of data availability are, so that readers can know whether work might be verified independently in principle when assessing its credibility.

Data access is only one part of transparency relevant to verification. Guidelines by political scientists have highlighted what they term “production transparency” and “analytic transparency,” with the former involving documentation of data collection and preparation processes and the latter involving documentation of steps by which findings are derived from data. Authors doing secondary analysis may believe the details in their articles are sufficient, but experience from efforts to reproduce results may suggest otherwise. For example, Breznau (2015) reports that Brooks and Manza (2006:436) declined requests for information to help reproduce findings from publicly available data that they published in the *American Sociological Review*, but Breznau’s efforts based on his reading of the article led to estimates that were not especially close to the original.

Page limits once severely restricted the information authors might be expected to provide about their research procedures, but online supplements and the ability to link to online sources have eliminated technical barriers. In *Science*, a group of mostly behavioral and social scientists published a set of “Transparency and Openness Guidelines” that journals could adopt, tiered into three levels of stringency (Nosek et al. 2015). Whether fields do so, or press for increased transparency in other ways, is now a matter for social and political settlement within epistemic cultures.

As with replication rules more broadly, no particular movement to improve transparency in sociology is known to us at present. The longtime editor of one of sociology's flagship journals offered one of the few arguments we have seen for how transparency research procedures could be scientifically bad: "There may be something very useful in not having clarity in data conventions and even in having code mistakes hidden from inspection. Maybe those kinds of things are necessary to produce (false) findings that lead us up out of local optima... in the great state space of truth" (Abbott 2007). External pressures for more transparent and reproducible research practices in sociology may increase, especially for interdisciplinary-minded sociologists, as expectations in neighboring social sciences increase. Fortunately, available guidance for doing reproducible research has increased enormously in recent years (Long 2009; see also collection of course syllabi at <https://osf.io/vkhbt/>).

Standards that make work more verifiable are sometimes opposed on the grounds that insufficient resources exist to actually do the verification (e.g., Abbott 2007). Transparency advocates stress that much of the scientific benefit of transparent practice is likely realized even if only rarely anybody actually verifies results. As importantly, the practices involved in making work transparent to others may improve the work by pressing researchers to be clear about the rationale for their own practices. It may also make work more efficient as insofar as documentation that makes work transparent to others is likely the best way of ensuring it will be able to be understood by collaborators and, perhaps most urgently, by oneself later in the long gestation that characterizes many research projects (Bowers 2011).

Robustness

Researchers may re-examine data from a published result to perform "sensitivity analyses" or "robustness tests" beyond those reported. This work seeks to assess whether key results are consistently observed across alternative ways the original analysis could have been done. Freese and Powell (2001) find that any of a series plausible alternative analyses to those presented by Kanazawa (2001) yield results that are greatly attenuated and no longer statistically significant. Young (2009) demonstrates that Barro and

McCleary's (2003) findings about a nation's religiosity and its economic growth change considerably under fairly small changes to model specification.

An abiding concern is that published findings represent a best-case scenario among all the arbitrary and debatable decisions made over the course of analyzing data. This might be due to “p-hacking,” where a researcher runs different analyses until they find support for their preferred hypothesis (Simonsohn et al. 2014), or to “HARKing” in which hypotheses presented as a priori were actually developed after results were known (Kerr 1998). Even when authors of an original study report extensive robustness tests of their own, one might still worry that researchers are biased toward reporting such tests when they strengthen findings but not when they weaken them. None of this necessarily requires any conscious deceptive motive, but the end result is that reported p values may greatly overstate the prospects that the observed finding was simply due to chance.

Various proposals have been introduced to provide ways of reducing latitude for p-hacking or HARKing (e.g., Wagenmakers et al. 2012; Nosek et al. 2012). Journals have been encouraged to articulate specific standards for what researchers are expected to disclose about the analyses they have done—for example, whether outliers were dropped or subgroup analyses were performed—and journals may even provide a checklist in which authors must affirm they have followed the journals expectations regarding disclosure (Eich 2014; Nosek et al. 2015). The website equator-network.org collects guidelines that have been offered for health research. Reporting guidelines do not prevent deception, but making disclosure a matter of rule draws a line where otherwise they could be disagreement about convention (Frow 2012; Freese and Peterson 2016).

In experimental fields, there has also been movement toward “pre-registering” studies by publicly archiving design and analyses plans beforehand, which allow authors to demonstrate that the hypotheses really were a priori and that the presented analyses were not cherry-picked based on results (Gonzales and Cunningham 2015). Pre-registration allows researchers to allay one type of potential distrust in their accounts of the research process. It does not displace the need for trust entirely: the biggest fraud scandal

of recent years in political science, notably, involved a field experiment that was pre-registered, but the ensuing data was simply fabricated (LaCour and Green 2014).

Re-analysis of original data may also be done when someone believes they have a better way of analyzing data than what was done in the original study. Killewald and Bearak (2010) argue that Budig and Hodges's (2010) study of the motherhood penalty should have used unconditional quantile regression instead of conditional quantile regression, and that this change dramatically changes the study's main result (see also rejoinder by Budig and Hodges [2014]). Lall (forthcoming) re-analyzes every comparative political economy study published in two journals over a five-year period and finds that key results become non-significant when multiple imputation for missing values is used instead of listwise deletion. Of course, such work is only possible when data from the original study are available to other investigators, and so providing a base to explore possible improvements in technique provides another justification for encouraging researchers to make data and code available.

Repeatability

Of course, replication involves more than reanalyzing old data. Concerns about a potential "replication crisis" in experimental psychology and biomedical research have focused on whether reportedly effects can be also observed by other scientists collecting *new* data. For example, Bem (2011) published a paper in a leading social psychology journal reporting a series of experiments providing apparent evidence of the paranormal phenomenon of precognition. Ritchie, Wiseman, and French (2012) report three separate efforts to follow the protocols of one of these experiments exactly; none of these found any evidence of the reported effect.

While this exchange may seem to exemplify ideals of scientific replication, it was actually a key episode in raising alarms for many that the replication conventions in social psychology may be problematic (e.g., LeBel and Peters 2011). For observers who flatly reject paranormal claims, that standard methods were used to generate evidence for precognition served as a high-profile demonstration

of the vulnerability of standard methods to false positives. More dramatically, the Ritchie et al. study was also desk-rejected by the journal that published the Bem article, on the grounds that it did not publish replication studies. To critics, the combined lesson was that ludicrous findings could find a prominent home in psychology journals while refutations of those findings could not.

With the Bem study, critics-of-the-critics can wonder whether replications really were entirely faithful to the original study in key details; the phenomenon could depend on some condition no one understands yet. Moreover, Ritchie et al. seemed to undertake the replication because they were skeptical. This skepticism could have led them to be careless about details—thinking that they couldn't matter anyway, since they believed the phenomenon to be non-existent. Moreover, those trying to replicate original findings may have more to gain from a failed replication than a successful one.

Recent activism in social psychology has led to journal policy changes that are explicitly more open to publishing direct replication studies. Yet, formal changes to replication rules do not resolve the difficulties in interpreting attempted replications. Indeed, when a large-scale effort to repeat 100 experiments published in three psychology journals found that only 39% successfully replicated (Open Science Collaboration 2015), a different team re-analyzed the Open Science Collaboration data and argued that claims of a “replication crisis” were greatly exaggerated (Gilbert et al. 2016a), which led in turn to another round of disagreements (Anderson et al. 2016; Gilbert et al. 2016b).

Social psychology has been involved in numerous initiatives to strengthen the credibility of replication studies (Brandt et al. 2014). Pre-registration, noted earlier, provides details about protocols in advance, and may address concerns about researchers making decisions as data are being analyzed that are biased toward their favored results. Additionally, researchers involved in the original study may be asked to evaluate protocols before the new data are collected (see Kahneman 2014 on “replication etiquette). The pre-registered experiment can be made available for other researchers to join if they agreed to adhere to the protocols, allowing for a much larger and presumptively more generalizable size than the original study. Journals may be asked to accept these multi-lab replications before data have

been collected, eliminating any perception that publication depends on the replication study's results (Simons et al. 2014).

While these initiatives have generated enthusiasm from activists, our familiarity with science studies' lessons about replication makes us unsurprised that these incentives have not been as powerful for generating consensus as some advocates may have hoped. For example, a recent large-scale replication effort involving 23 labs and over 2000 subjects failed to replicate results of a previously published experiment finding evidence of the phenomenon of ego depletion (Hagger and Chatzisarantis 2016). Afterward, two social psychologists closely associated with ego depletion argued that "we understood our approval [of the project's design] to mean 'Sure, go ahead' and not 'Yes, that's a definitive test of the phenomenon we've been studying all these years.'" (Baumeister and Vohs 2016: 574).

In their view, ego depletion has been demonstrated by a range of different experiments in different domains, and the failure to replicate this one experiment, no matter how comprehensive, does not undermine all these results. In social psychology this has sometimes been characterized in terms of a debate over the value of "conceptual replications," which are deliberately different studies that are intended to test an existing theory in a novel way. Conceptual replications overcome the problem of replications being devalued as noncreative work by allowing researchers to develop imaginative ways to test existing theories. But conceptual replications have been criticized because, when they fail, they are difficult to publish because of the chronic ambiguity about whether they actually speak to the theory in question. When conceptual replications are distrusted, failed direct replications may be taken as having much graver implications for a larger theory than might appear logically warranted.

Generalizability

Of the different forms of replication we discuss, repeatability likely accords best with commonplace understandings of replication. Nevertheless, across quantitative social science discussions

of replication, repeatability per se is noticeably de-emphasized relative to the re-examinations of original data described earlier, on the one hand, and inquiries directed to the generalizability of findings on the other (see, e.g., Wilson et al. 1973; Lucas et al. 2013). The former is illustrated the first paper to describe sociology as having a “replication problem,” and Wilson et al. (p. 141-149) there present replication specifically with respect to generalizability: “It is not the task of the scientist to uncover truths specific to the particular experimental setting he employs. A key test of the generality of scientific knowledge is that of replicability.”

The emphasis on generalizability is also evidenced in the frequency with which social scientists use “replication” to refer to standardizing measures or holding constant parts of a research design, in order to facilitate comparison. For example, the General Social Survey keeps the same wording of core items from year after year, and the International Social Survey Programme coordinates the asking of the same survey questions across member nations. Panel studies ask respondents the same questions, and studies of social continuity or change have asked visited the same sites in different eras. All these follow, in different ways, a dictum we have heard attributed to Otis Dudley Duncan, “If you want to measure change, don’t change the measure.” The last *Annual Review of Sociology* paper ostensibly about replication focused predominantly on such issues, particularly in the context of collecting longitudinal data (Bahr, Caplow, and Chadwick 1983).

We do not think that this reflects unusual epistemology by quantitative social science so much as the typical sorts of data used. Most quantitative social science uses data that were not collected by the authors of the paper, and were instead assembled as large-scale public goods for a research community. The nature of these datasets is that often they are uniquely suited for the questions posed using them—their unique potential is part of what justifies the expense of their existence—and, even when the same question can be put to multiple such datasets, the datasets differ in ways that are prominent to the understanding of any findings.

Even when social scientists do collect their own data, these data often exhaust the narrow scope of their results in a way that is not compatible with repeatability as subsequent research goal. To

illustrate, Smith and Papachristos (2016) present findings from data they assembled of 3,000 persons connected with organized crime in the Chicago area from 1900 to 1950. Regarding the replicability of the conclusions, we might imagine that someone could challenge their results by asserting problems with their analysis or with the data, and the latter could even involve someone re-doing some of the work involving primary sources. We can also imagine someone challenging the conclusions using data from organized crime elsewhere. What is *not* possible is trying to see if the results are observed again by repeating their research with an independent sample of 3,000 other organized crime figures in Chicago in the same period.

Of course, the examples of repeatability from psychology in the preceding section are also not studying the same population at the same time. Occasionally these differences come to the fore when understanding differences in replication (Schnall 2014), but we suspect sociologists may be surprised by how much they do not. Our own impression is that, when psychologists are faced with divergent results from studies asking the same question, they tend to first consider how the *designs* of the two studies are different, whereas sociologists in the same situation tend to first consider how the *samples* for the two studies differ. Experimental psychologists regularly use deliberately non-representative samples that are only interesting because they are presented as broadly illuminating about generalized human psychological processes; sociology focuses more on population samples or historical cases for which results might be taken as instructive even if they do not generalize farther than their target population.

Connecting particular findings to general conclusions is often done by reference to an animating theory that is supported by the original study's results but also asserts broader application. When Walker and Cohen write that "every general sociological proposition is both true and false" (Walker and Cohen 1985: 288), they mean is that general propositions are true under some conditions and false under others, a position elaborated by Cartwright (1999). To the extent this is true, what empirical work tests is not so much propositions as such but ideas about the scope in which the propositions apply.

For example, the “fundamental cause” theory of disease proposes that the inverse relationship between SES and health emerges as a result of differential returns to advancing knowledge about how to prevent and treat disease. Evidence presented for the theory includes evidence of SES and health for causes of death that have become preventable and the absence of such relationships when causes are absent (Link and Phelan 1995; Lutfey and Freese 2005). Meanwhile, strong socioeconomic gradients in diseases that preceded accurate understanding of their causes or prevention, such as what existed in Britain during early cholera outbreaks (e.g., Johnson 2006), would indicate some need to amend thinking.

When replication is formulated in terms of evaluating generalizability, the advantage of designs that introduce differences seems plain (Lucas et al. 2013). Showing that a relationship between SES and diabetes mortality is repeatedly observed, even if clarifying that specific empirical matter, does little to advance a broader understanding of the scope over which SES negatively affects health. In the case of well-defined scope conditions, diverse findings may be said to be replicated when consistent with the expected scope. In practice, scope conditions are often not well-defined, and, regardless, findings inconsistent with existing notions about the scope of a proposition seem common and provide one way new findings offer innovations on existing understandings.

The chronic ambiguity is whether explanations revised to accommodate new results constitute a legitimate advance, or are simply ad hoc explanations that mask deeper problems with the validity of some of the studies in question. Ultimately, social scientists must resolve such questions on a case-by-case basis. A general lesson, however, is that worries about the basic validity of original findings often leads to efforts to repeat studies. To the extent this avenue is often foreclosed in social science, it is understandable that one might see greater concern at least for seeing if findings are verifiable, and, beyond this, one might suppose such enterprises would have ongoing difficulty resolving various ad hoc explanations of inconsistent results.

REPLICATION IN QUALITATIVE SOCIAL SCIENCE

The flip side of saying replication is essential to science is that “replication” may be viewed as a shibboleth of scientism in social inquiry. The editor of a political science journal that focuses on qualitative research in political science recently refused to adopt transparency standards proposed for the discipline because he viewed them as embodying a “neo-positivism” in which political science becomes “not a never-ending contest between perspectives on politics but instead about the veridical understanding of the world as a set of objective processes” (Isaac 2016: 275-276). For qualitative sociology, Abrahamson and Dohan (2015: 273) note that methodological discussions around replication are complicated because qualitative researchers “do not necessarily share basic epistemological assumptions about the research enterprise with either their quantitative colleagues or one another.”

Qualitative social science is often presented as firstly an interpretive enterprise. In sociology, it is typically concerned with issues of motivation and meaning that give the *verstehen* approach its lasting appeal. It may be that “replication” is simply the wrong language to apply to qualitative studies. Attempts to make qualitative research more scientific through intermediate steps like rigorous “coding” might only obscure unavoidable interpretive work (Biernacki 2012). Those who disagree will need to be more explicit about what replication means in fields defined by thick forms of evidence.

Sometimes cited as a counterpart to failed “replication” in qualitative research is the famous episode in which Derek Freeman (1983, 1999) disputed Margaret Mead’s classic anthropological study of adolescent sexuality in Samoa. But this work did not involve Freeman trying to deliberately “redo” Mead’s fieldwork, but rather he drew on his own expertise on Samoa and revisited some of Mead’s sources. More to the point, Freeman’s critique largely circumvents the difficult issues surrounding interpretive inquiry by accusing Mead of simply being wrong on objectively available facts. While this bears some affinities to “verification” as discussed above, the better analogue to any sort of replication may be “fact-checking” in journalism: that is, determining whether details presented as matters of objective fact can be independently verified. Fieldwork methods often involve considerations of

confidentiality that prompt researchers to use pseudonyms and even change details to prevent identification of parties or settings, which can strongly limit possibilities for fact-checking.

Anonymization practices are routine in fieldwork studies, but doing so has also been occasionally questioned for the lack of researcher accountability (Duneier 1999; Murphy and Jerolmack 2016). Desmond (2016) recently offers the intriguing alternative of hiring an independent, named fact-checker who was given access to otherwise confidential materials. Besides limiting fact-checking, obscuring details may constrain the ability of work to be extended by others or to be subjected to alternative interpretations. Murphy and Jerolmack (2016) note an example in which Bosk (2008) reflected on his classic ethnography of medical errors from three decades earlier. Because he changed the gender of the only female medical resident in the group he studied, others were not able to consider whether gender had shaped interactions.

Once one moves from “brute facts” (Taylor 1985) to matters of interpretation, the premise of fact-checking becomes murkier. Of course, there is the question of whether others who had access to the same materials would have interpreted them in the same way. For archival research, for instance, questions of whether interpretations of primary sources are idiosyncratic are longstanding and potentially addressed by other experts having access to the same archives. An intriguing recent is “active citation,” in which authors provide far more detail to citations, including very specific hypertext links to the original materials for specific claims (Moravcsik 2010).

Abrahamson and Dohan (2015) offer a new graphical display, the “ethnoarray,” which, they argue, may aid ethnographic replications by providing more explicit systems for coding evidence. Hammersley (1997) is encouraged by potential for growing archives of ethnographic data to aid in the assessment of claims and support re-analysis, but he cautions against the idea that ethnographic findings can be “audited” in a way equivalent to quantitative datasets. He argues that fieldnotes are open to different interpretations and are a part of a process that largely occurs within the original writer’s head.

Importantly, the usual premise in qualitative research is that the understanding of any piece of primary material might not be correctly understood solely from that material. Rather, it requires a larger understanding of the setting that the researcher has gained over the extensive course of their project. As Collins (1998) puts it: “Interview extracts illustrate what analysts understand to be going on as a result of their experiences in the field and are used to convey this understanding to a less expert audience; they are not thought of as data.”

Small (2009) provides an influential argument for the centrality of the analyst in qualitative research. He advocates an iterative procedure in which qualitative researchers “replicate” observations through a number of distinct cases (following language of Yin [2002]). After each case, the research questions evolve to become more sophisticated. The process ends once the researcher subjectively feels “saturated” and believes that additional sites will yield little new information.

Although Small uses the language of replication, the procedure he advocates is very different from the forms of evaluative replication that we describe above. As Small points out, qualitative data collection often changes as one’s questions evolve. Unchanging and ubiquitous aspects of a site disappear from fieldnotes as the analyst’s interests become refined. Subtle phenomena emerge and become central topics. The presence or absence of some phenomena in fieldnotes is meaningless because they are not an unbiased record of a site. Unlike most forms of quantitative data, fieldnotes lose much of their value when divorced from the notetaker.

CONCLUSION

The foregoing has sought to bring together sociological research on replication in science with recent discussions about improving replication practices in social science. It should be clear that none of the four forms of replication in quantitative sociology completely overcome the issues of ambiguities and interpretation that sociologists of science have observed. We conclude by offering a typology of the

ambiguities that different sorts of the different types of replication described above confront, and then by reflecting briefly on what we consider the primary normative takeaway for replication policies in our own discipline of sociology.

The Routine Ambiguities

We presented replication in quantitative social science as divisible into four goals—verifiability, robustness, repeatability, and generalizability. In each case, the interpretation of results from replication efforts confront routine, often inexorable, ambiguities which often serve to structure professional debates around contentious replications (Table 3).

[TABLE 3 ABOUT HERE]

Verifiability gives rise to the ambiguity of intentionality. Because verification is directed so specifically to the details of another work, interpretation of verification problems often raise questions about the authors of the original study's intentions. When failures are due to problems in the original studies, should the failure be interpreted as a problem of authorial honesty, competence, or care—that is, as something that should have reputational consequences? Or does it just reflect the occasional bugginess that researchers recognize as a pragmatic reality of their craft?

Tests of robustness can result in an ambiguity of judgment. When reanalyzing old data under new specifications, the debate often turns on whether certain decisions were appropriate, and whether the consequences of these decisions ultimately matter for the conclusions that are drawn. This raises the potential for expert disagreement on whether such choices represent incorrect practice or a defensible exercise of professional discretion.

When a replication attempts collect new data, the ambiguities center on Collins' discussion of similarity and difference. The ambiguity of similarity can emerge from tests of repeatability. Attempts to

repeat an outcome with new data can be met with charges that the data diverge from the original in ways substantial enough to undermine any equivalence. On the other hand, the attempt to generalize a finding can produce the ambiguity of difference since replicating in new environments or using new methods can weaken the link between the replication and target study.

Replication policy in sociology

As sociologists, the most striking thing in reviewing recent developments in social science replication is how much all our neighbors seem to be talking and doing about improving replicability. Reading economists, it is hard not to connect their relatively strict replication culture with their sense of importance: shouldn't a field that has the ear of policy-makers do work that is available for critical inspection by others? The potential for a gloomy circle ensues, in which sociology would be more concerned with replication and transparency if it was more influential, but unwillingness to keep current on these issues prevents it from being more influential. In any case, the integrative and interdisciplinary ambitions of many sociologists are obviously hindered by the field's inertness on these issues despite the growing sense in nearby disciplines that they are vital to ensuring research integrity.

At the same time, many sociologists pride ourselves on the way that being the most undisciplined social science promotes creativity. Discussions elsewhere have ended up with proposals for top-down regulation of research practice. We strongly suspect such regulation works best when the regulating entity (e.g., a journal, a professional organization) shares an epistemic culture. Cases in psychology and political science demonstrate that arguments about "raising standards" can be easily interpreted as treating particular types of research as the prototype and marginalizing others. Even when one epistemic culture is powerful enough to impose its regulatory ideas on others, doing so prioritizes one group's methodological progress over the many values of disciplinary pluralism.

By the same token, epistemic communities need to be secure and generous enough not to impede efforts at improvements by other communities with whom they share a discipline. Code and data sharing

guidelines for quantitative sociology should move forward even if there is no counterpart for qualitative sociology.

One possibility are the badges that have been adopted by some journals in psychology and political science, which allow papers that follow certain guidelines about certain practices—for instance, open data or preregistration—to advertise that they have done so. Different communities could, if they wish, develop different badges reflecting what they regard as important. Pluralistic disciplines flourish when their different constituents both work together when helpful yet give one another the latitude to strengthen separately. Changing ideas about replication almost certainly present a case where the different epistemic cultures of sociology need to help one another find their own ways.

Table 1: Replication Success and Similarity on Interpretation

	Similar	Different	
Success	Strengthens narrow (not broad) interpretation	Strengthens broad (and narrow) interpretation	← Objection: Differences too great for comparison
Failure	Undermines narrow (and broad) interpretation	Undermines broad (not narrow) interpretation	

↑
Objection:
Critical differences
introduced

Table 2: Forms of Replication in Quantitative Social Science

	Similar	Different
Old Data	Verifiability	Robustness
New Data	Repeatability	Generalization

Table 3: The Four Routine Ambiguities

	Similar	Different
Old Data	Ambiguity of Intentionality	Ambiguity of Judgment
New Data	Ambiguity of Similarity	Ambiguity of Difference

REFERENCES

- Abbott A. 2007. Notes on Replication. *Sociolog. Methods & Res.* 36(2):210-9
- Abramson CM, Dohan D. 2015. Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data. *Sociolog. Methodol.* 45(1):272-319
- Anderson CJ, Bahnik S, Barnett-Cowan M, Bosco FA, Chandler J, et al. 2016. Response to Comment on “Estimating the Reproducibility of Psychological Science”. *Science.* 351(6277):1037
- Ashmore M. 1989. *The Reflexive Thesis: Wrighting Sociology of Scientific Knowledge.* Chicago, IL: University of Chicago Press
- Bahr HM, Caplow T, Chadwick BA. 1983. Middletown III: Problems of Replication, Longitudinal Measurement, and Triangulation. *Annu. Rev. of Sociol.* 9:243-64
- Barro, RJ. and McCleary, RM. 2003. Religion and Economic Growth across Countries. *Am. Sociol. Rev.* 68(5):760-81
- Baumeister RF, Bratslavsky E, Muraven M, Tice DM. 1998. Ego Depletion: Is the Active Self a Limited Resource? *J. of Person. and Soc. Psychol.* 74(5):1252-65
- Baumeister RF, Vohs KD. 2016. Misguided Effort with Elusive Implications. *Perspect. on Psychol. Sci.* 11(4):574-5
- Bem D. 2011. Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect. *J. of Person. and Soc. Psychol.* 100:407-25
- Bernanke BS. 2004. Editorial Statement. *Am. Econ. Rev.* 94(**):404
- Biernacki R. 2012. *Reinventing Evidence in Social Inquiry: Decoding Facts and Variables.* New York: Palgrave Macmillan

- Bosk CL. 2008. *What Would You Do?: Juggling Bioethics and Ethnography*. Chicago: University of Chicago Press
- Bowers J. 2011. Six Steps to a Better Relationship with Your Future Self. *The Political Methodol.* 18(2):2-8
- Brandt, MJ, IJzerman H, Dijksterhuis A, Farach, FJ, Geller, J, Giner-Sorolla R, ... van't Veer A. (2014). The replication recipe: What makes for a convincing replication? *J Exp Soc Psychol.* 50: 217–224.
- Breznau N. 2015. The Missing Main Effect of Welfare State Regimes: A Replication of ‘Social Policy Responsiveness in Developed Democracies’ by Brooks and Manza. *Sociol. Sci.* 2:420-41
- Brooks C, Manza J. 2006. Social Policy Responsiveness in Developed Democracies. *Am. Sociol. Rev.* 71(3):474-94
- Brown AN, Cameron DB, Wood BDK. 2014. Quality evidence for policymaking: I’ll Believe it When I See the Replication. *J. of Dev. Eff.* 6(3):215-35
- Budig MJ, Hodges MJ. 2010. Differences in Disadvantage: Variation in the Motherhood Penalty across White Women’s Earnings Distributions. *Am. Sociol. Rev.* 75(5):705-28
- Budig MJ, Hodges MJ.. 2014. Statistical models and empirical evidence for differences in the motherhood penalty across the earnings distribution. *Am Sociol. Rev.* 79(2):358-64
- Burawoy M. 2003. Revisits: An Outline of a Theory of Reflexive Ethnography. *Am. Sociol. Rev.* 68(5):645-79
- Cartwright N. 1999. *The Dappled World: A Study in the Boundaries of Science*. Cambridge, UK: Cambridge University Press

- Clemens MA. 2015. The Meaning of Failed Replications: A Review and Proposal. *J. of Econ. Surv.* <http://ftp.iza.org/dp9000.pdf>
- Cole S. 1992. *Making Science: Between Nature and Society*. Cambridge, MA: Harvard University Press
- Collins HM, Evans R. 2002. The Third Wave of Science Studies: Studies of Expertise and Experience. *Soc. Stud. of Sci.* 32(2):235-96
- Collins HM. 1974. The TEA Set: Tacit Knowledge and Scientific Networks. *Sci. Stud.* 4(2):165–86
- Collins HM. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press
- Collins HM. 1991. The Meaning of Replication and the Sciences of Economics. *Hist. of Political Econ.* 23(1):123-42
- Collins HM. 1998. The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves. *Am. J. of Sociol.* 104(2):293-338
- Collins HM. 2001. “Tacit Knowledge, Trust and the Q of Sapphire.” *Soc. Stud. of Sci.* 31(1):71-85
- Collins HM. 2010. *Tacit and Explicit Knowledge*. Chicago, IL: University of Chicago Press
- Collins HM. 2016. Reproducibility of Experiments: Experimenters’ Regress, Statistical Uncertainty Principle, and the Replication Imperative. In *Reproducibility: Principles, Problems, and Prospects*, ed. Harald Atmanspacher, Sabine Maasen, pp. 65-82. New York: John Wiley & Sons, Inc.
- COS (Center for Open Science). 2013. “Badges to Acknowledge Open Practices.” Retrieved from <https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/>

- Daston L, Galison P. 2010. *Objectivity*. New York: Zone Books
- Daston, Lorraine. 1992. Objectivity and the Escape from Perspective. *Soc. Stud. of Sci.* 22(4):597-618
- Desmond M. 2016. *Evicted: Poverty and Profit in the American City*. New York: Crown
- Doing P. 2004. 'Lab Hands' and the 'Scarlet O': Epistemic Politics and (Scientific) Labor. *Soc. Stud. of Sci.* 34(3):299–323
- Duneier M. 1999. *Sidewalk*. Chicago, IL: University of Chicago Press
- Edwards P, Mayernik MS, Batcheller AL, Bowker GC, Borgman CL. 2011. Science Friction: Data, Metadata, and Collaboration. *Soc. Stud. of Sci.* 41(5):667-90
- Eich, Eric. 2014. "Business Not As Usual." *Psychol Sci.* 25(1):3-6
- Finkel EJ, Eastwick PW, Reis HT. 2015. Best Research Practices in Psychology: Illustrating Epistemological and Pragmatic Considerations with the Case of Relationship Science. *J. of Person. and Soc. Psychol.* 108(2):275-97
- Freeman D. 1999. *The Fateful Hoaxing of Margaret Mead: A Historical Analysis of Her Samoan Research*. Boulder, CO: Westview Press
- Freeman D. 1983. *Margaret Mead and Samoa*. Cambridge, MA: Harvard University Press.
- Freese J, Peterson D. 2016. The Emergence of Forensic Objectivity. osf.io/2ft8x
- Freese J, Powell B. 2001. Making love out of nothing at all? Null findings and the Trivers-Willard hypothesis. *Am. J. of Sociol.* 106(6):1776-86
- Fujimura JH. 1988. The Molecular Biological Bandwagon in Cancer Research: Where Social Worlds Meet. *Soc. Probl.* 35(3):261–83
- Galison P, Stump DJ. 1996. *The Disunity of Science: Boundaries, Contexts and Power*. Palo Alto, CA: Stanford University Press

- Gilbert DT, King G, Pettigrew S, Wilson TD. 2016a. Comment on Estimating the Reproducibility of Psychological Science. *Science*. 351(6277):1037
- Gilbert DT, King G, Pettigrew S, Wilson TD. 2016b. A Response to the Reply to Our Technical Comment on ‘Estimating the Reproducibility of Psychological Science.’ http://projects.iq.harvard.edu/files/psychology-replications/files/gkpw_response_to_osc_rebutal.pdf?m=1457031863
- Goldberg, PK. 2015. Report of the Editor: *Am. Econ. Rev.* 105(5):698-710
- Gonzales, JE, Cunningham CA. 2015. “The promise of pre-registration in psychological research.” *Psychological Science Agenda*. Retrieved from <http://www.apa.org/science/about/psa/2015/08/pre-registration.aspx>.
- Guo G, Yi L, Tianji C, Hongyu W, Greg D. 2015. "Peer influence, genetic propensity, and binge drinking: a natural experiment and a replication." *Am. J. of Sociol.* 121(3):914-54
- Hacking I. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. New York: Cambridge University Press
- Hacking I. 1996. The Disunities of the Sciences. In *The Disunity of Science: Boundaries, Contexts and Power*, ed. P Galison, D Stump, pp. 37–74. Palo Alto, CA: Stanford University Press
- Hagger MS, Chatzisarantis NLD. 2016. A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspect. on Psychol. Sci.* 11(4):546-73
- Hammersley, M. 1997. Qualitative Data Archiving: Some Reflections on its Prospects and Problems. *Sociology*, 31(1):131-42
- Herndon T, Ash M, Pollin R. 2014. Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff. *Camb. J. of Econ.* 38(2):257-79

- Hughes V. 2014. "Simple Steps Aim to Solve Science's 'Reproducibility Problem.'" Retrieved from <https://spectrumnews.org/news/simple-steps-aim-to-solve-sciences-reproducibility-problem/>
- Hunter A. 1983. The Gold Coast and the Slum Revisited: Paradoxes in Replication Research and the Study of Social Change. *Urban Life*. 11(4):461-76
- Isaac JC. 2015. For a more Public Political Science. *Perspect. on Politics* 13(2):269-83
- Jacoby, WG. 2016. "AJPS to Award COS Open Practice Badges." Retrieved <https://ajps.org/2016/05/10/ajps-to-award-cos-open-practice-badges/>
- Johnson S. 2006. *The Ghost Map: The Story of London's Most Terrifying Epidemic*. London: Penguin
- Kahneman, Daniel. 2014. A new etiquette for replication. *Soc Psychol* 45(4): 310-311.
- Kanazawa, S. 2001. Why we Love our Children. *Am. J. of Sociol.* 106(6):1761-76.
- Kennefick D. 2000. Star Crushing: Theoretical Practice and the Theoreticians' Regress. *Soc. Stud. of Sci.* 30(1):5-40
- Kerr, Norbert L.. 1998. HARKing: Hypothesizing after the results are known. *Pers Soc Psychol B* 2(3): 196-217.
- Killewald A, Bearak J. 2010. Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression. *Am. Sociol. Rev.* 79(2):350-7
- King, G. 1995. Replication, Replication. *PS: Political Sc. & Politics* 28(9):444-52
- Kitcher, Philip. 1995. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press
- Knorr Cetina, K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press

- LaCour MJ, Green DP. 2014. When contact changes minds: An experiment on transmission of support for gay equality. *Science*. 346(6215):1366
- Lall R. 2016. How Multiple Imputation Makes a Difference. *Political Analysis*.
http://scholar.harvard.edu/files/ranjitlall/files/pa_website_version_1.pdf?platform=hootsuite
- Leahey E. 2008. Overseeing Research Practice: The Case of Data Editing. *Sci. Technol. & Hum. Values*. 33(5):605-30
- LeBel EP, Peters KR. 2011. Fearing the Future of Empirical Psychology: Bem's (2011) Evidence of PSI as a Case Study of Deficiencies in Modal Research Practice." *Rev. of Gen. Psychol.* 15(4):371-9
- Lieberman ES. 2010. Bridging the Qualitative-Quantitative Divide: Best Practices in the Development of Historically Oriented Replication Databases. *Annu. Rev. of Political Sci.* 13:37-59
- Link BG, Phelan JC. 1995. Social Conditions as Fundamental Causes of Disease. *J. of Health and Soc. Behav.* 35:80-94
- Long JS. 2009. The Workflow of Data Analysis Using Stata. *Stata Press*
- Longo DL, Drazen JM. 2016. Editorial: Data Sharing. *New Engl. J. of Med.* 374:276-7
- Lucas, Jeffrey W, Kevin Morrell, and Marek Posard. 2013. Considerations on the "Replication Problem" in Sociology. *The Am. Sociol.* 44:217-32
- Lutfey K, Freese J. 2005. Toward some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in the Routine Clinic Visit for Diabetes. *Am. J. of Sociol.* 110:1326-72
- Mack, RW. 1951. The Need for Replication Research in Sociology. *Am. Sociol. Rev.* 16(1):93-4.

- McCullough BD, McGeary KA, Harrison TD. 2006. Lessons from the Journal of Money, Credit, and Banking Archive. *J. of Money, Credit, and Bank.* 38(4):1093-107
- McPherson M, Smith-Lovin L, Brashears ME. 2006. Social Isolation in America: Changes in Core Discussion Networks over Two Decades. *Am. Sociol. Rev.* 71(3):353-75
- McPherson M, Smith-Lovin L, Brashears ME. 2008a. Erratum: Social Isolation in America: Changes in Core Discussion Networks over Two Decades. *Am. Sociol. Rev.* 73(6):1022
- Mead M. 1928. *Coming of Age in Samoa*. New York: William Morrow & Company
- Meier KJ. Replication: A View from the Streets *PS: Political Sci. & Politics.* 28(3):456-9
- Merton RK. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.
- Moravcsik A. Active Citation: A Precondition for Replicable Qualitative Research. *PS: Political Sci. & Politics.* 43(1):29-35
- Mulkay M. 1984. The Scientist Talks Back: A One-Act Play, with a Moral, about Replication and Reflexivity in Sociology. *Soc. Stud. of Sci.* 14(2):265-82
- Murphy A, Jerolmack C. 2016. Ethnographic masking in an era of data transparency. <https://contexts.org/blog/ethnographic-masking-in-an-era-of-data-transparency-2/>
- Nature*. 2014. Journals unite for reproducibility. 515:7
- Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, et al. 2015. Promoting an Open Research Culture. *Science.* 348(6242):1422-5
- Nosek BA, Spies JR, Motyl M. 2012. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth of Publishability. *Perspect Psychol Sci,* 7(6): 615-631.
- Open Science Collaboration. 2015. Estimating the Reproducibility of Psychological Science. *Science.* 349:943

- Panofsky A. 2014. *Misbehaving Science: Controversy and the Development of Behavior Genetics*. Chicago, IL: University of Chicago Press
- Parish E, Bloom T, Godlee F. 2015. Statins for People at Low Risk. *BMJ*. 351:h3908
- Peterson D. 2015. All that is Solid: Bench-Building at the Frontiers of Two Experimental Sciences. *Am. Sociol. Rev.* 80(6):1201-25
- Peterson RR. 1996a. A Re-Evaluation of the Economic Consequences of Divorce. *Am. Sociol. Rev.* 61(6):528-36
- Peterson RR. 1996b. Statistical Errors, Faulty Conclusions, Misguided Policy: Reply to Weitzman. *Am. Sociol. Rev.* 61(6):539-40
- Pickering A. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago, IL: University of Chicago Press
- Pinch TJ. 1979. Normal Explanations of the Paranormal: The Demarcation Problem and Fraud in Parapsychology. *Soc. Stud. of Sci.* 9(3):329-48
- Political Science Replication. 2015. *Leading journal verifies articles before replication--so far, all replications failed.*
<https://politicalsciencereplication.wordpress.com/2015/05/04/leading-journal-verifies-articles-before-publication-so-far-all-replications-failed/>
- Popper K. 1992. *The Logic of Scientific Discovery*. New York: Routledge
- Porter TM. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press
- Reinhart CM, Rogoff KS. 2010. Growth in a Time of Debt. *Am. Econ. Rev.* 100(5):573-8
- Ritchie SJ, Wiseman R, French CC. 2012. Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect. *PLoS ONE*. 7(3):e33423

- Rose AM. 1953. Generalizations in the Social Sciences. *Am. J. of Sociol.* 59(1):49-58
- Rothstein J. 2007. Does Competition Among Public Schools Benefit Students and Taxpayers?
Comment. *Am. Econ. Rev.* 97(5):2026-37
- Schnall S. 2014. Clean Data: Statistical Artifacts Wash Out Replication Efforts. *Soc. Psychol.*
45(4):315-7
- Shapin S. 1989. The Invisible Technician. *Am. Sci.* 77(6):554–63
- Shapin S. 2010. *Never Pure: Historical Studies of Science as if It Was Produced by People with Bodies, Situated in Time, Space, Culture, and Society, and Struggling for Credibility and Authority.* Baltimore, MD: Johns Hopkins University Press
- Simons DJ, Holcombe AO, and Spellman BA. 2014. “An introduction to registered replication reports at *Perspectives on Psychological Science.*” *Perspect Psychol Sci* 9(5): 552-555.
- Simonsohn U, Nelson LD & Simmons, JP. 2014. “P-Curve: A Key to the File-Drawer Problem.” *J Exp Soc Psychol*, 143(2), 534-547.
- Small ML. 2009. How Many Cases do I Need? On Science and the Logic of Case Selection in Field-Based Research. *Ethnography.* 10(1):5-38
- Smith CM, Papachristos AV. 2016. Trust Thy Crooked Neighbor: Multiplexity in Chicago Organized Crime Networks. *Am. Sociol. Rev.* 81(4):644-67
- Sripada C, Kessler D, Jonides J. 2014. Methylphenidate Blocks Effort Induced Depletion of Regulatory Control in Healthy Controls. *Psychol. Sci.* 25(6):1227-34
- Taylor, C. 1985. *Philosophical Papers: Volume 1, Human Agency and Language.* Cambridge, UK: Cambridge University Press
- Tischler HL. 2007. *Introduction to Sociology.* Ninth Edition. Belmont, California: Wadsworth

- Travis GDL. 1981. Replicating Replication? Aspects of the Social Construction of Learning in Planarian Worms. *Soc. Stud. of Sci.* 11(1):11-32.
- Wagenmakers EJ, Wetzels R, Borsboom D, van der Maas HLJ & Kievit, RA 2012. "An Agenda for Purely Confirmatory Research." *Perspect Psychol Sci.* 7(6): 632-638.
- Walker HA, Cohen BP. 1985. Scope Statements: Imperatives for Evaluating Theory. *Am. Sociol. Rev.* 50(3):288-301.
- Weitzman L. 1985. *The Divorce Revolution: The Unexpected Social and Economic Consequences for Women and Children in America.* New York: The Free Press
- Wicherts JM, Borsboom D, Kats J, Molenaar D. 2006. The Poor Availability of Psychological Research Data for Reanalysis. *Am. Psychol.* 61(7):726–8
- Wilson FD, Smoke GL, Martin JD. 1973. The Replication Problem in Sociology: A Report and a Suggestion. *Sociol. Inq.* 43(2):141-9
- Yearly S. 2005. *Making Sense of Science: Understanding the Social Study of Science.* London: SAGE Publications
- Yin, R. 2002. *Case Study Research.* Thousand Oaks, CA: SAGE
- Young C. 2009. Model Uncertainty in Sociological Research: An Application to Religion and Economic Growth. *Am. Sociol. Rev.* 74(6):380-97
- Young C, Horvath A. 2015. Sociologists Need to be Better at Replication.
<https://orgtheory.wordpress.com/2015/08/11/sociologists-need-to-be-better-at-replication-a-guest-post-by--young/>
- Zuckerman, H. 1977. Deviant Behavior and Social Control in Science. In *Deviance and Social Change*, ed. Edward Sagarin, pp. 87-138. Beverly Hills, CA: SAGE Publications