ANNUAL REVIEWS

Annual Review of Criminology

Replication in Criminology and the Social Sciences

William Alex Pridemore,[1] Matthew C. Makel,[2] and Jonathan A. Plucker[3]

[1]School of Criminal Justice, University at Albany, State University of New York, Albany, New York 12222, USA; email: pridemore@albany.edu

[2]Talent Identification Program, Duke University, Durham, North Carolina 27701, USA

[3]School of Education, Johns Hopkins University, Baltimore, Maryland 21209, USA

Annu. Rev. Criminol. 2018. 1:19–38

First published as a Review in Advance on October 25, 2017

The *Annual Review of Criminology* is online at criminol.annualreviews.org

https://doi.org/10.1146/annurev-criminol-032317-091849

Keywords

replication, reproducibility, open science, research methodology, content analysis

Abstract

Replication is a hallmark of science. In recent years, some medical sciences and behavioral sciences struggled with what came to be known as replication crises. As a field, criminology has yet to address formally the threats to our evidence base that might be posed by large-scale and systematic replication attempts, although it is likely we would face challenges similar to those experienced by other disciplines. In this review, we outline the basics of replication, summarize reproducibility problems found in other fields, undertake an original analysis of the amount and nature of replication studies appearing in criminology journals, and consider how criminology can begin to assess more formally the robustness of our knowledge through encouraging a culture of replication.

INTRODUCTION

The public is losing its trust in scientific findings. Although this may reflect a general public questioning of expertise (Nichols 2017), a more fundamental factor may be at play: Research quality is often lacking. Nearly universal prescriptions to floss our teeth and take our vitamins have come under public scrutiny due to questionable empirical support (Donn 2016, Guallar et al. 2013). Such highly visible examples from daily life can lead the public to overgeneralize and doubt the state of the scientific enterprise and the evidence behind all policy recommendations. This mistrust is especially important to criminology. Unlike flossing teeth and taking vitamins, we study topics that are not only highly visible and with widespread societal consequences but also that are often the source of contentious public and political debate. Nearly every citizen feels strongly about issues like capital punishment, gun control, and the causes and consequences of crime. If the public, practitioners, and policy makers do not trust the findings of criminologists in this highly politicized environment, then who can be trusted?

Recent examinations of research quality in several fields revealed that doubt in empirical findings is not unfounded. Published results are often difficult or impossible to replicate, and a range of questionable research practices (QRPs) have been discovered. Begley & Ellis (2012) were able to replicate the findings of only 11% of highly cited preclinical cancer trials. Using data on psychology articles over nearly 30 years, Nuijten et al. (2016) found more than half contained inconsistencies in reporting p-values that often led to incorrect conclusions about statistical significance. O'Boyle et al. (2014) labeled "the chrysalis effect" the tendency of dissertation results in the field of management to become more likely to support hypotheses when eventually published in journals. Unsupported dissertation hypotheses tended to disappear or change direction when published, while newly supported hypotheses appeared (the latter may occur more generally and sometimes be facilitated by a publishing process that leads authors to state findings in terms of hypotheses that were generated as part of the manuscript review process). The general public finds QRPs and outright scientific fraud to be highly objectionable and possibly criminal (Pickett & Roche 2017), suggesting that growing knowledge of QRP prevalence and low replication rates can erode public trust in research.

The concerns about replication that have swept through so many other disciplines have largely passed over criminology. This presents meaningful threats to the field. Historically, the methods employed in studies published in our journals on average have been lower on the evidence hierarchy than those in medicine and psychology, two domains in the midst of replication crises. Further, policies or interventions in areas like policing, punishment, and crime prevention hold the potential to do widespread harm and can be an enormous drain on the public treasury. It is difficult to argue for evidence-based crime and justice policy if criminologists are unable to assure the public of the trustworthiness of their evidence (Sampson et al. 2013, Tittle 2004, Wellford 2010).

Widespread false positives represent a central aspect of the replication problem. The vast majority of published studies report statistically significant results, even in the biomedical sciences where research methods are usually more rigorous than in the social sciences (Chavalarias et al. 2016). Studies of humans that apply social and behavioral methods are more than three times as likely to report positive results as studies of nonbiological material in physics and chemistry (Fanelli 2010). An extremely small fraction of published studies are replications, so nearly all test novel hypotheses. One must be skeptical that such a large proportion of novel hypotheses are correct, and given the hierarchy of the sciences this is especially true for social sciences like criminology. Some sizeable fraction of these novel findings—much greater than the 5% false-positive rate conventionally accepted when setting statistical significance at 0.05—must, in reality, be nonsignificant. Ioannidis (2005a) went further, arguing most published research findings are

false. The scientific record will be obscured by the prevalence of false positives as long as replication research remains so rare.

WHAT IS REPLICATION?

Replication is the use of subsequent studies to purposefully assess the same questions as original studies. Replication is the main arbiter by which research findings are adjudicated because it protects against false positives. It has been referred to as the Supreme Court of science (Collins 1985). Ideally, evidence for or against a hypothesis is accumulated over time by scholars independent of each other. In criminology, we are often interested in the implications of our research for policy and harm-reduction interventions, so we also want to know if original findings hold for other individuals, neighborhoods, cultures, institutions, and times. Novel findings may be compelling, but if they are not robust their contribution to the scientific record are limited. "While currently there is unilateral emphasis on 'first' discoveries," stated Ioannidis, "there should be as much interest on replication of discoveries" (quoted in de Bono 2005). Despite being a cornerstone of the scientific process, replication often is not covered in detail in textbooks (see Schmidt 2017), including criminology textbooks.

Replication is variously defined (Christakis & Zimmerman 2013) but generally falls into two main categories: direct and conceptual. Direct replication is when a scholar attempts to use the same type of data and methods as the original study to assess the veracity of its findings. This is possible only when the original publication (or communication with the author) provided sufficient information for the study to be repeated. Conceptual replication is when scholars attempt to test the original idea using different data or methods, purposely altering some component to test the generalizability of the original results. External and operational replications fall into this category. External replication is using data collected from new samples, different settings, or different times to determine whether original findings hold in new contexts (Kline 2004). This type of replication otherwise follows the original design and methods as closely as possible. Operational replication is a new study in which the primary hypothesis remains the same but updated or superior data, designs, or methods are employed. Freese & Peterson (2017) generally use the same language but organize quantitative replications in the social sciences into four categories based on the goals of verifiability, robustness, repeatability, and generalizability. Figure 1 in Condon et al. (2017) provides a simple helpful way to distinguish between the terms used in replication research.

Multiple replications of all studies and total reproducibility of all research are not possible and probably not desirable (Freedman et al. 2015), as this would create tremendous cost burdens and limit the diversity of phenomena criminologists study. Below, we provide recommendations for prioritizing what merits replication.

Why Is Replication Important?

Replication is important because it strengthens confidence in the scientific record. It helps turn an initial observation into more widely accepted knowledge. Replication plays a key role in verifying scientific hypotheses and results and in assuring findings are generalizable beyond the specific circumstances of a particular study (Schmidt 2009). Replications act like a kidney, filtering out inaccurate or irreproducible findings that otherwise pollute our understanding of a phenomenon (Makel & Plucker 2014b). The systematic accumulation of knowledge is central to the scientific method, and replicability is a key element in this process. Replications that confirm original results increase certainty in the existence and size of an effect.

A recent analysis of studies in three psychology journals showed replications of even high-powered designs were able to reproduce only approximately one-third of the original findings (Open Sci. Collab. 2015). There is little reason to believe similar replications of criminology studies would fare better. The retraction rate of publications is also steadily rising (Van Noorden 2011; see also **http://retractionwatch.com**). If the vast majority of publications present novel and positive findings, and if a large majority of these findings do not replicate, then privileging novelty has meaningful detrimental effects on the evidence base. When highly cited studies are not subjected to replication or their findings are not reproducible, it damages our understanding of how the world operates, hampers theoretical development, and can lead to harm and substantial wasted time, effort, and money following false leads.

Why Is Replication Important in Criminology?

Replication is important in criminology because our evidence base is tenuous, many of our topics are highly politicized, and translation of our findings can produce harm if they are incorrect. Criminology is not immune to the replication crisis and on average our studies may be at greater risk of failure to replicate than those in many other fields. The cumulative irreproducibility rate (Freedman et al. 2015) in criminology is likely due mainly to study design, data analysis, and reporting of results. As a social and behavioral science, criminology is ranked low on the hierarchy of sciences, and the methods typical of our field are low on the evidence hierarchy. Further, the behaviors of humans, groups of humans, and institutions like those that make up the criminal justice system are complex, and criminologists frequently lack complete data on our phenomena of interest and must often rely on indirect measures of theoretical concepts. In the face of these limitations, putting stock in one or a small set of publications on a topic is unwise and replication is essential.

Research findings on crime and justice may have widespread real-world effects on individuals, groups, and societies via policy, practice, and manipulation of public sentiment by politicians and other advocates. A key goal of much of the research on crime and justice is harm reduction of some sort, but there are likely more ways to produce harm due to unintended consequences than to reduce it. The quality of the scientific record takes on added importance in this context, and one way to achieve greater assurance in the accuracy of our evidence base is to increase replication research. If our genuine goal is to discover the truth and sometimes use that information to reduce harm, promote justice, and make criminal justice systems more efficient, then scrutinizing the rigor of our accumulated knowledge via replication should not be perceived as a criticism of or threat to individual scholars or specific studies but instead simply part of the normal practice of science.

What Makes a Successful Replication?

What is required to ensure a replication truly replicates the original study? When can we state with confidence the replication reproduces the same results as the original study?

A replication study can never be the same as the original in every respect. In general, the expectation is the replication is as similar as possible to the original study in its hypotheses and methods. If a minor alteration in method leads to very different results, however, then the original finding is unstable and thus likely of limited use for researchers or practitioners because much of the value of research findings lies in their expected ability to help predict subsequent outcomes under similar conditions. If minor changes in method produce different results, this must be made clear, and some argue empirical papers should include a section that helps readers understand the constraints and the expected breadth of generality of the finding (Simons et al. 2017). At the same time, sometimes sameness in design is not desired. When the goal is to determine if the findings confirm the

effect under different conditions and thus provide generalizability, purposefully changing some aspect of the original study can be beneficial. Schmidt (2009) describes in detail this combination of similarity and difference in replication research (see also Freese & Peterson 2017).

Replicating a study as closely as possible can require substantial and close communication with the original authors to acquire data and materials and/or to obtain precise information on methods used during and decisions made at various stages of the research process. Some journals (e.g., *Psychological Science*) are changing word limits to exclude method and analysis sections to encourage researchers to include as much relevant information as possible. This is important because authors are not always willing or able to discuss their methods. Brandt et al. (2014) provide a recipe for what makes a convincing replication.

A replication is generally considered successful if it reproduces the same results as the original study, although the precise meaning of "the same results" remains up for debate. Two standard approaches to determining replication success are discovering whether the effect size in the replication is significantly different from zero and discovering whether it is significantly different from the original effect size (Simonsohn 2015). There are notable limitations to both methods. Simonsohn (2015) summarized these limitations and proposed what he terms a detectability approach that combines hypothesis testing and estimating effect size (see also Schwarz & Clore 2016, Simonsohn 2016), with attention to increasing sample size in the replication to provide greater power to detect an effect. In social and behavioral sciences, there is rarely a critical test that provides a definitive answer (Lieberson & Lynn 2002), and the same is true when assessing the success of replications. Instead, like in a jury trial, we weigh evidence for and against an idea. Thus, a study that fails to replicate an original finding is only one piece of evidence, which in itself means multiple replications may be required to form a body of evidence large enough to make an informed judgment.

REPLICATION RESEARCH

There are many well-known examples of replication research in the physical and social sciences. Some of these include successful replications. An early classic example was the work of Robert Boyle, Robert Hooke, and Christiaan Huygens to demonstrate the latter's discovery of the anomalous suspension of water inside a vacuum pump [Shapin & Schaffer 1985; see, especially, "Chapter VI, Replication and its Troubles: Air-Pumps in the 1660s"]. Burger (2009) and Dolínski et al. (2017) replicated the main findings from Milgram's (1963) study of obedience. Other well-known examples include unsuccessful replications, like those of cold fusion as reported by Fleischmann & Pons (1989; see Petrasso et al. 1989), and precognition as reported by Bem (2011; see Ritchie et al. 2012). These examples illustrate some common positive and negative aspects of replication. Huygens traveled from Amsterdam to London to work closely with and advise Hooke, who had until then been unable to reproduce the finding. Fleischmann & Pons (1989), however, communicated poorly or not at all with skeptics attempting to replicate their results. And Ritchie et al.'s (2012) replication was rejected by the journal that published Bem's original study.

The Replication Crisis Replicates

Replication has always been part of science, and scientists have always lamented its infrequency. In recent years, however, scholars increasingly began to scrutinize systematically and in writing the fragile research findings in field after field. Replication crises in medicine and psychology received the most attention, but other fields like advertising, biology, computer science, genomics, library sciences, marketing, medicine, political science, public health, and sociology also became more cognizant of reproducibility (for a brief review, see Makel & Plucker 2014a). Recognition of the

problem often led to a crisis of confidence in the accumulated evidence base. A survey by *Nature* of researchers primarily from the physical and life sciences found 52% of respondents believed there was a significant reproducibility crisis and 90% believed there was at least a slight crisis (Baker 2016). This feeling was neither abstract nor driven wholly by the attention the topic is receiving in journals and the media: Nearly three-quarters of respondents had tried and failed to reproduce another scholar's findings, more than half had tried and failed to reproduce their own findings.

Ioannidis (2005b) found that only 44% of results from the highly cited medical studies he examined were successfully replicated by later studies, and effect sizes were stronger in initial analyses relative to replications. The latter is a familiar phenomenon (see also Ioannidis 2008) commonly labeled the decline effect. Begley & Ellis (2012) were able to reproduce the findings of just 6 of 53 (11%) published studies in cancer biology, and Prinz et al. (2011) successfully replicated only 35% of the published medical findings they examined. These examples may be uncomfortable to criminologists, most of whom likely believe the methods employed in medical research are on the whole more rigorous than those in the social sciences.

Psychology has received the most replication attention among the behavioral and social sciences. A large-scale effort to reproduce 100 published findings from three leading psychology journals found that although 97% of the original findings had been statistically significant, only 36% of replication attempts were significant, and mean effect sizes in the latter were half those in the former (Open Sci. Collab. 2015). There have also been several "many labs" attempts to replicate previously published psychology findings in different research labs around the world (see **https://osf.io/89vqh/**). Economists have also considered reproducibility. One fundamental aspect of the field is understanding individuals' preferences for goods and experiences. Replication research raises questions about prior findings on this topic and more generally describes why one should be skeptical of new results and why nonreplication in the field is to be expected (Maniadis et al. 2014). In an example from political science, Lall (2016) examined the impact of using multiple imputation instead of listwise deletion in studies of comparative and international political economy. Using nearly every such study in two key journals over five years he was unable to reproduce nearly half the key findings. The experiences of our social science colleagues with reproducibility also may make some criminologists uncomfortable.

A Few Examples with Implications for Criminology

Criminology has not systematically taken part in this evaluation of and introspection about the reproducibility of its knowledge base. There are, however, instances of unsuccessful replications with various implications for the field. Failure to replicate does not automatically mean the original finding was untrue and the replication finding was true. Still, when independent researchers cannot replicate results, confidence in the effect declines and there are lessons to be learned.

Ego depletion is a highly influential theory in social psychology that has received substantial academic and popular coverage over the past 20 years (Baumeister et al. 1998). The argument is that, like one's physical strength, self-control is a finite resource. Thus, after mental exertion, including acts of choice and of self-control itself, the ability to exercise self-control is depleted (Baumeister 2002). Reduced ability to function, impulsive behavior, and poor choices ensue. This idea has direct relevance to criminology, including obviously for self-control theory. Hagger et al. (2016) carried out a multi-lab preregistered replication of ego-depletion findings using a standardized protocol. The meta-analysis of results from their 23 different labs and 2,141 subjects showed no evidence of the ego-depletion effect.

Saperstein & Penner (2010, p. 92) found that incarceration influenced how one's race, regardless of phenotype, was perceived by others and by themselves, concluding "...race is not a fixed

characteristic of individuals but is flexibly and continually negotiated in everyday interactions." This finding received widespread attention and has several provocative individual- and structural-level consequences beyond those related to punishment. Hannon & Defina (2016) replicated and reexamined the study and were unable to reproduce the original findings, concluding incarceration does not strip people of racial privilege.

In an attempt to examine the impact of poverty on national homicide rates, Pridemore (2008, 2011) first set out to replicate the findings on other structural covariates from approximately four dozen cross-national homicide studies using data collected by the original authors. He had to abort this exercise after contacting authors of more than half the studies because only two provided their original data. Pridemore (2011) was able to reproduce the key findings of these two studies. He then attempted to replicate the findings from another cross-national homicide study, at the time very recently published in a prominent journal, whose authors were unable to provide their data. He collected and analyzed the data using the description of the methods provided in the original publication. He was unable to reproduce original findings for the key covariates, and the two sets of estimates were not close.

A well-documented example in criminology is from the Minneapolis Domestic Violence Experiment. Sherman & Berk (1984) found that relative to domestic violence offenders who were given counseling or temporarily removed from the household, those who were arrested were less likely to reoffend. This was a landmark study demonstrating the use of experimental methods in the field of criminology, and it had a substantial influence on police practice in responding to domestic violence incidents. Several of the replications of the experiment, however, were unable to reproduce the original deterrent effect or found a weaker effect (see Berk et al. 1992, Sherman et al. 1991).

Limitations of Replication Research

There are limitations of replication research and its utility. Although replication research has the potential to reveal results due to QRPs or that are ephemeral rather than enduring, it is not a cure for all research problems. As mentioned above, multiple replications and total reproducibility of each study are not possible or desirable, and questions remain about what constitutes a replication and when we can state with confidence that a replication reproduces the same results as the original study. There are several other limitations.

Research that is replicable can still be wrong (Leek & Peng 2015). Thus, truth may be overestimated because replications come to the same erroneous conclusions as original studies. In the presence of publication bias, an inaccurate finding may appear well-replicated despite numerous but unpublished failed attempts to replicate it. Due to widespread QRPs, even internal conceptual replication—which involves using the original sample in various ways, such as splitting it in two for cross-validation purposes or drawing a large number of random samples from it to estimate the precision of the parameters of interest—does not appear to increase success of independent replications (Kunert 2016). Many factors, including QRPs, threaten the validity of results in original and replication study pairs. Replicating what are in reality invalid findings generates substantial confusion because it leads to the creation of stylized facts that are actually untrue.[1] In this way, replication can produce opacity and mistrust in the scientific record instead of clarification.

[1]A related concern is multiple tests of the same flawed data, which highlights the importance of independent replications. If all or many studies of a phenomenon employ the same data set and produce the same finding, then we do not have many replications but many re-analyses, with the limits to external validity this entails. Although the ability of independent groups to reproduce the same finding with the same data is of fundamental importance, our confidence grows further when independent data yield the same finding, and this gets to the mission of replication.

Conceptual replications are not as rare as the meta-research reported here suggests. Thus, truth may be underestimated because we think there are fewer replications than really exist. Schmidt (2017) called these disguised conceptual replications. Criminologists, for example, are especially interested in the generalizability of an effect to other people, groups, places, and times. There are countless articles of this type of research that are not framed as replications and that do not use some form of the word "replicate" when published, which hides them from meta-research on replications and potentially from researchers looking for validation of findings.

Some believe the replication crises may be overblown. In psychology, for example, Gilbert et al. (2016) argued the widespread pessimism in the wake of the Open Science Collaboration (OSC) (2015) study, which was able to replicate less than half of the 100 original studies, is unwarranted due to error, power, and bias. First, the OSC effort used substantially different methods in many of the replications, which introduced random error. Second, the OSC effort attempted to replicate each study only once, reducing power and underestimating the true replication rate. Third, there may have been bias toward failure. The OSC study asked original authors if they endorsed the fidelity of the protocol prior to the replication being carried out, and 60% of endorsed protocols but only 15% of unendorsed protocols resulted in successful replications. Gilbert et al. (2016) concluded the OSC data in fact suggested reproducibility in psychological science is high not low (but see Anderson et al. 2016 for a persuasive counter to these critiques). Similarly, using the OSC data, Patil et al. (2016) found that 77% of OSC replications were within a 95% prediction interval calculated using the original effect size, which returns to our earlier discussion of the disagreement over what is considered a successful replication.

Some analyses found successful replication is common: 67% in education journals (Makel & Plucker 2014a), 79% in psychology (Makel et al. 2012), and between 66% (Lemons et al. 2016) and 80% in special education journals (Makel et al. 2016) and 67% and 78% for laboratory experiments in economics (Camerer et al. 2016). However, these self-identified successes were based simply on whether the replicating authors stated that their results were similar to the original findings, not on any formal calculation. Finally, the Many Labs Project took a different approach compared to the OSC. It repeated 16 classic psychology studies in 35–36 different labs and then pooled the data, successfully replicating 85% of the original findings (Klein et al. 2014).

Nonreplication

Often there is simply a lack of replication research on a particular topic. Sometimes replication research on a topic exists but fails to replicate the original findings. We consider both alternatives here.

Lack of replication research. The very small amount of replication research relative to its importance is problematic. One cause of this, especially in criminology and related fields, is a set of data-related challenges. Authors may be unwilling or unable to provide their data to replicators (Wicherts et al. 2006, Young & Horvath 2015; for an example in criminology, see Pridemore 2011), and no matter how clear the presentation of methods in the publication, it cannot contain a list of all decisions made along the way. Criminologists also often rely on secondary data. Thus, even when analyses are perfectly executed, we are at the mercy of an imperfect data-generation process and the inevitable human errors, such as spreadsheet coding and transcription errors, that occur when manipulating it (Herndon et al. 2014) and that might lead to irreproducible results. When criminologists do undertake original data collection, the nature or scope of such projects can making repeating them highly improbable, for example the large-scale multigenerational Rochester Youth Development Study or ethnographic fieldwork like that of Hureau (2016; see

also Hureau & Braga 2018) examining the nature of urban violence and the market for illicit guns in high-risk networks (for a brief introduction to replication issues in qualitative social science, see Freese & Peterson 2017).

The main cause of the lack of replication research is an epistemic culture that discourages it. Cultural impediments to undertaking and publishing replications derive from structure and practice (Freese & Peterson 2017). Grant funding and space in leading journals are scarce resources. Universities incentivize research production, including grants and top-tier publications. Funding agencies, journal editors, and journal reviewers reward innovation and discourage replication. Journals also favor positive findings (except in the case of positive replications, in which case the likely response is "Thanks, but we already knew that."), further reducing the likelihood of publication when the original result is not reproduced. Informally, one's approach to replication is usually learned from mentors early on, often only subtly and in contexts not directly related to it. Doing the work required that allows others to reproduce research—such as keeping a log of every data and analysis decision, retaining and carefully labeling all code, and making all this publicly available in a clear protocol—takes time that might be spent producing more original research. As a replicator, there are no formal rules about what merits replication, and asking scholars for data, code, and other information can be perceived as a threat (Kahneman 2014). The structural factors create competition in an environment of scarce resources that privileges novelty over truth, thereby devaluing replicable research and replications. Informal practices adapt to that environment and help transmit the epistemic culture's values about replication. Thus, the scarcity of replication research seems to come down to incentives. Actions that foster and facilitate replication help the field but do not necessarily help (and may even harm) the individual researcher. If incentives can be better aligned, actions will follow.

Reasons for unsuccessful replication. When replication research does occur, why does it often fail to reproduce original results?[2] Finding the truth is not easy, so some of these failures simply reflect the normal practice of science. Replication is precisely playing its role when it helps separate discovery from distraction. What remains unclear, however, is the extent to which unsuccessful replications are due to the inherent difficulty of conducting science and how much they are due to original findings being biased in some way. As with all complex phenomena, false positives have many interrelated causes.

First, there is bias toward statistical significance. Publication bias is well established, appears to be getting worse (Fanelli 2012), and is unlikely to be the result of improving methodological rigor (Van Noorden 2011). Publication bias is so strong in the social sciences that negative findings are not only less likely to be published but even less likely to be written up (Franco et al. 2014). Pigott et al. (2013) studied this outcome reporting bias in education research by comparing dissertations with their subsequent published versions. Of the 621 dissertations they examined, 79 had been published as a journal article. Dissertation results that were nonsignificant were 30% less likely to appear in the journal article than dissertation results that were significant. Publication bias selects for false positives and inflates effect sizes of published studies (Ioannidis 2008) and thus increases the risk of unsuccessful replication. However, the bias toward statistical significance is not caused solely by publication bias (Makel & Plucker 2014a). The competitive environment discussed above, for example, likely exacerbates the problem. If one is rewarded for publishing, and publishing requires statistical significance, then false positives will be common.

[2]When a finding is not reproduced, the replication may be at fault. Similarly, the original finding may be true but a conceptual replication reveals the association is contextual and not universal. Our attention here, however, is on the original false positive.

Second, scholars possess a tremendous number of "researcher degrees of freedom" when collecting and analyzing data and communicating results. Simmons et al. (2011) demonstrated this freedom by showing impossible results (in their example, individuals becoming younger after listening to a song) can be cultivated with sufficient statistical contortion. To use an aphorism familiar to criminologists, any good prosecutor can convince a grand jury to indict a ham sandwich. Although this liberty can be used unscrupulously, it is not necessarily so. Each decision can be made in good faith, but the countless decisions made in the research process add up and can influence the outcome. This flexibility is one key realm in research where self-deception occurs (see Nuzzo 2015). To avoid becoming a cargo-cult science when exercising these researcher degrees of freedom, we must remember "[t]he first principle is that you must not fool yourself—and you are the easiest person to fool" (Feynman & Leighton 1985; the quote was originally in Feynman's 1974 commencement address to Caltech, which can be found at **http://calteches.library.caltech.edu/51/2/CargoCult.htm**).

Third, limitations in data and method can produce findings that are not reproducible. It is generally accepted there is an inverse correlation between the likelihood of detecting an association and the strength of the method used to detect it. Ioannidis (2005a) concluded that results were less likely to be true when sample and effect sizes were small and when potential conflicts of interest, bias, and flexibility in design were high. Although cancer research was the focus of Begley & Ellis' (2012) review, criminologists will recognize many of the common limitations of irreproducible results they mentioned: lack of controls, several data-related problems, not repeating one's own analyses to check results, and incorrect use of statistical methods. As we show below for criminology, another regular finding is that independent replications (i.e., those that do not share authors with the original publication) are less likely to be successful (Makel et al. 2012).

Fourth, researchers engage in questionable practices that degrade research quality. These can overlap with researcher degrees of freedom and with limitations in data and method, but here our emphasis is on purposeful acts. John et al. (2012) surveyed more than 2,000 research psychologists, asking if they had ever engaged in what the authors deemed QRPs like falsifying data or results, selectively reporting only significant results, claiming to have predicted unexpected findings, arbitrarily excluding outliers, failing to report all dependent variables, and deciding if more data needed to be collected after looking at initial results. Although less than 1% of respondents admitted falsifying data or results, nearly half admitted selectively reporting only studies that worked. More than half admitted not reporting all dependent variables and peeking at results to determine whether to collect more data. These findings make it obvious a replication undertaken by the same authors may not control for QRPs or fraud. A recent study by Brodeur et al. (2016) revealed empirical evidence of one form of the QRP popularly known as p-hacking. Examining all p-values reported in three prestigious economics journals between 2005 and 2011, the authors found heaping of p-values just under 0.05, with far fewer than expected p-values between 0.10 and 0.25. This suggests authors selectively report specifications that are significant and/or initially find "nearly significant" results then manipulate model specification until $p < 0.05$. With such high self-admission rates and empirical evidence of p-hacking in leading journals, it is clear that QRPs are not rare, that they could play a substantial role in false positives, and that they undermine the accuracy of the scientific record.

THE AMOUNT OF REPLICATION RESEARCH IN CRIMINOLOGY JOURNALS

In this section, we present original research that estimated how many replications are being published in criminology journals. We also examine what types of replications (direct versus

conceptual) were published, whether the authors determined that the replications were successful, whether any original authors were involved in the replication and whether this resulted in a higher chance of successful replication, and which journal (i.e., same as original article or new journal) published the replication. McNeely & Warner (2015) undertook a similar original study of replication in criminology, although they limited their analysis to five leading journals over five years (2006–2010) and focused on somewhat different questions than we addressed here.

Method

In autumn 2015, we used the Web of Knowledge search engine to search the entire population history (through 2014) of all 52 journals in the Criminology & Penology category of the *ISI Web of Knowledge Journal Citation Reports Social Sciences Edition* to identify all published articles and all articles containing any form of the term replicat* in the text (using the "Topic" option in Web of Knowledge). This replicates methods used previously in psychology, education, and special education research (Makel & Plucker 2014a; Makel et al. 2012, 2016). There were 40,020 total articles published in these journals through 2014. We removed the 745 articles from *Deviance et Societe*, *Recht & Psychiatrie*, and *Revija za Kriminalistiko in Kriminologijo* because they were not published in English. We searched the remaining 39,275 articles for the term replicat*, which yielded 283 articles.

We coded these 283 articles according to (*a*) whether new data were collected to replicate a previous finding, (*b*) whether the answer to the first question was yes, and if so then what type of replication was carried out (direct, conceptual, or a mixture of both), (*c*) whether the analyses were considered by the authors to have successfully replicated original results, and (*d*) whether any authors on the replication were also authors on the original research being replicated.

Results and Discussion

Of the 283 articles containing replicat*, we coded 178 as actual replications of previously published findings. Of the articles that used the term replicat* that were not replications, 41% suggested the need for replication. Other uses of the term replicat* included citing replication studies and suggestions for replication of policies.

The 178 actual replications represented only 0.45% of the 39,275 criminology articles we searched. This is similar to the rate found in special education (estimated at 0.41% by Lemons et al. 2016 and 0.5% by Makel et al. 2016) and is higher than in education (0.13% in Makel & Plucker 2014a) but is less than half that in psychology (1.07% in Makel et al. 2012). In their analyses of five leading criminology journals from 2006 to 2010, McNeely & Warner (2015) found 2.34% of articles were replications. This may reflect a difference in how McNeely & Warner defined and coded replications relative to our approach. It might also reflect important differences in publishing replications in the five leading journals they examined relative to all 52 journals in *ISI* we examined, differences in recent years (they examined 2006–2010, we analyzed the entire publication history as reported by *ISI*), or both.

Table 1 shows replication outcome by replication type and authorship type. Nearly 75% of all replications were self-identified by replicators to have successfully replicated the original finding, just over 10% failed to replicate, and approximately 15% reported mixed findings. This success rate is similar to that found by Makel and colleagues (2012, 2014a, 2016) in studies of special education (80%) and psychology (79%), and a bit higher than they found in education (67%).

More than 80% of replication studies were conceptual replications, 16% were direct replications, and 3% had aspects of both. We found no evidence replication success differed significantly

Table 1 Replication outcomes by replication type and authorship in criminology journals

Replication type	Replication outcome			
	Success	Failure	Mixed	Total
Direct	24 (82.8%)	2 (6.9%)	3 (10.3%)	29 (16.3%)
Conceptual	107 (74.3%)	17 (11.8%)	20 (13.9%)	144 (80.9%)
Mixed	1 (20.0%)	0	4 (80.0%)	5 (2.8%)
Total	132 (74.2%)	19 (10.7%)	27 (15.2%)	178 (100%)
Authorship				
Author overlap	79 (77.4%)	5 (4.9%)	18 (17.6%)	102 (57.3%)
Unique authors	53 (68.7%)	14 (18.4%)	9 (11.8%)	76 (42.7%)

between direct and conceptual replications [$\chi^2(2) = 0.99$, $p = 0.611$, Cramer's $V = 0.075$]. There were too few mixed replications (five) to make a statistical comparison of success.

Replicating authors had been part of the original research team in 57% of replications, and these cases were approximately 13% more likely than cases without author overlap to result in successful replication [$\chi^2(2) = 8.77$, $p = 0.012$, Cramer's $V = 0.22$], although both groups self-identified as successfully replicating original findings the majority of the time (78% for overlap, 69% for nonoverlap). Close communication with original authors is often important with replications, and authorship overlap could help ensure fidelity of the replication that, in turn, helps replication success. However, if the finding is so fragile it requires authorship overlap, then it is problematic for the generalizability and usefulness of the result.

Approximately 45% of replications were published in the same journal as the original study it was replicating, whereas 55% were published either in a different journal or were replicating a finding from a book or a report that was not published in a journal. This provides some evidence that criminology may be following a form of the "Pottery Barn rule," where journals that publish findings are also responsible for publishing replication attempts (Srivastava 2012).

ENCOURAGING A CULTURE OF REPLICATION

The challenge of reproducibility and the threat it poses to our evidence base will eventually reach criminology. Damage to the field could be significant because of the implications of our research for policy and practice, the cost and potential harm of interventions, the strongly partisan response elicited by many of the phenomena we study, and the visibility and integration of crime and justice in everyday life. Some responses to this threat from other fields translate well to criminology (see McBee et al. 2017, Munafo et al. 2017, Spellman et al. 2017), and we encourage proactive consideration (Makel & Plucker 2017). In this section, we discuss what merits replication, individual and systemic strategies that encourage a culture of reproducible research and replication, and a few existing proreplication efforts in criminology.

What Merits Replication?

Even in a small field like criminology, a tremendous amount of research produces an overwhelming number of journal articles and books each year. Not all findings merit replication, and replication research requires substantial time and effort, so there must be a way to prioritize which studies are replicated.

Individuals possess various motivations for choosing to replicate previous research. Some scholars replicate their own work to bolster their confidence in their results. Some volunteer to replicate

the findings of a colleague in exchange for the same service. One new online platform to facilitate this type of mutual replication is called Study Swap (**https://osf.io/view/StudySwap/**). Some scholars want to be sure of the accumulated evidence on a particular topic before moving forward with their own research that will extend knowledge in the area. Others may simply doubt the outcome of a study—skepticism, after all, is a key ingredient of science—and so attempt to replicate its findings. Replication can also be a powerful tool for training graduate students (Grahe et al. 2012) and may lead to early publications for them.

More systematically, to protect the scientific record and to ensure the reputation of a field, we believe several types of findings should be prioritized for replication. First, highly cited findings should be replicated. There are many reasons research gets cited, and results often get reified as they become highly cited. Independent replication, not citation, should make for stylized facts. Second, findings likely to be translated to policy or practice should be replicated. Given the prevalence of publication bias, QRPs, and other limitations, independent replication should play a prominent role before widespread implementation. Third, findings with implications for theoretical foundations should be replicated. For example, conclusions about measuring theoretical constructs and testing key theoretical assumptions are important to replicate because they are used in all other tests of a theory. Fourth, findings that are regularly discussed in introductory textbooks should be replicated. Introductory textbooks are read by more people than advanced texts. If findings that have not been replicated are treated as fact in introductory textbooks, perhaps due to their novelty or because they are timely or popular topics, it misleads an amateur but broad audience. Finally, findings that receive significant attention should be replicated. Journals privilege novelty and we are all interested in intriguing new findings, but hot topics deserve scrutiny. This includes findings that receive media coverage. Fields do not have control over which results get selected by the media, but scholars have a responsibility to mention the necessity of independent replication of findings when speaking to the media about research.

Individual Strategies that Encourage a Culture of Replication

Individual scholars can do many things to encourage a culture of replication. The most important is simply to remember the overarching goal of research is to make a meaningful contribution to the scientific record. This helps scholars resist perverse incentives for positive findings. This in turn helps scholars avoid implementing QRPs that contribute to false positives, damage the evidence base, waste the money, time, and effort of others by creating false leads, and spawn replication crises. As one example, Wicherts et al. (2016) provided a comprehensive checklist for avoiding p-hacking. Scholars can further resist publication bias by writing up and trying to publish the null findings that are so important for advancing knowledge (Franco et al. 2014). Individuals might also consider conducting replications as a service to the discipline, like reviewing manuscripts (LeBel 2015). Both replication and manuscript review protect the scientific record, and although replication takes considerably more effort, the replicator gets credit for publications and the benefits that flow from them. Similarly, when serving as manuscript reviewers, scholars can support a replication culture by not automatically recommending rejection for replication studies.

Another critical step individuals can take to contribute to a culture of replication is undertaking replicable research (Stodden et al. 2016). Scott Long's (2009) book on the workflow of data analysis is a remarkable achievement to this end and an extremely valuable resource for social scientists. The Open Science Framework (2017) shares course syllabi that provide direction for carrying out reproducible research. Further, when writing up results, scholars should be as detailed as possible and strive to communicate information clearly and in a way that allows replication. This could include adopting the 21-word solution suggested by Simmons et al. (2012, p. 1): "We report how

we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study." Finally, a key element of reproducible research is willingness to share data, material, methods, and workflow with other scholars for replication purposes, whether in general (e.g., in an official registry or on one's own website) or in response to specific requests.

Systemic Strategies that Encourage a Culture of Replication

The negative outcomes of irreproducible findings come largely at the expense of the research community and the public (Freedman et al. 2015), including effects on the evidence base, a field's reputation, translation of research into policy and practice, and inefficient use of limited funding. Therefore, systemic initiatives to encourage a culture of replication and of open science methods more generally are required. These efforts must include professional organizations, universities, funding agencies, and especially journals.

Among the most important things an academic field and its professional organizations can do to instill positive scientific values is promote goals that advance knowledge rather than individual careers. Further, although innovation in theory and method is crucial to the research enterprise, novelty should not be encouraged at the expense of the reliability of the evidence base. An open research culture (Nosek et al. 2015) supports these goals. Professional organizations can support replication by including it in mission statements and codes of ethics, hosting training workshops, highlighting conference panels, and rewarding replicable research and replication in various ways. Of course, data transparency can raise matters of respondent confidentiality and proprietary use of data that scholars took years to collect. Some may also perceive that discussion of replication and open research practices elevates some types of research and demotes others. These are all important questions professional organizations can help navigate.

Universities and their departments can also support a culture of replication. University-wide offices of research can host colloquia on undertaking reproducible research and replication. Promotion and tenure committees should recognize the contributions to both research and service provided by high-quality replication studies that make meaningful contributions to the scientific record. Departments can offer graduate courses on replication, perhaps open to students from and with costs shared by multiple units to increase efficiency, and individual faculty members can integrate these topics into courses or discussions on research methods and the philosophy of science.

Funding agencies must play a major role in encouraging replication. Depending on the type of research called for in the request for proposals, these agencies can require open science methods for funding or they can favor proposals that include concrete plans to undertake replicable research and provide access to data and workflow.[3] Agencies should also set aside resources to carry out replications, including large-scale efforts like those of the Open Science Collaboration (2015) and the Many Labs Project (e.g., Klein et al. 2014). One recent example is a fund created by the Netherlands Organization for Scientific Research (2017) earmarked for replications of high impact studies that became "the basis for subsequent research or. . .assumed an important place in

[3]Norms on data sharing are not settled. On one hand, one might argue for incentives for investigators who spend years or even decades on original data collection. Reconciling this proprietary investment with demands for full access to data is difficult [for the field of criminology, see the exchange between Lauritsen (2005) and Thornberry (2005)]. One resolution has been an agreement with the funding agency to release data a certain number of years following their collection. On the other hand, this question reflects disciplinary differences. In some fields there is little expectation that data are proprietary but instead an expectation that data sharing is mandatory. Even in those fields, however, there is a struggle to determine how to provide appropriate credit to those who initially collected the data. The foundation of the argument for data sharing, and more generally for open science, is the same as that of replication: Scientists are fallible, and the scientific record benefits when we can more quickly sort wheat from chaff among research findings. We discuss data sharing in the context of the open science movement below.

education, policy forming or public debate." Funding agencies are often heavily involved in disseminating research findings to practitioners and policy makers, and they should act as gatekeepers to ensure findings that might be translatable have been replicated.

Finally, journals are critical to encouraging a culture of replication. Publication bias is to blame for a large proportion of false positives and is central to the replication crises. Researchers are human. They desire individual and professional gain, including job security, and respond to incentives. For those in research institutions, the key pathway to success is producing research, mainly publications. An obstacle to publishing even carefully executed research is publication bias. Facing this obstacle, scholars often resort to QRPs. Franco et al. (2014) found social scientists often do not write up null or negative results, and more than 60% of researchers surveyed by *Nature* (Baker 2016) stated that pressure to publish and selective publishing contribute to irreproducible research. If publication bias incentivizes QRPs and is a key source of false positives, then journals must work to recalibrate the evidence base by removing incentives for statistical significance, publishing more null findings, and publishing more replications, especially of key studies originally published in their own pages and especially if replications do not reproduce original findings.[4]

There are other ways for journals to promote replications. First, journals may exempt data and methods sections from word limits, allowing authors to describe procedures in detail. Now that they are online, journals may also allow authors to include supplemental information. Second, a small but growing number of journals have adopted guidelines that require sharing data, code, and workflow (Nosek et al. 2015), usually in an independent third-party archive. A few journals even require independent replication and verification of results based on that shared information. So far, however, journal guidelines for sharing data often are not followed in practice (Iqbal et al. 2016). Third, some journals have adopted the use of badges, which are a method for recognizing studies conducted using open science methods. These methods usually are not required to publish in the journals, but badges on the published article denote practices like preregistering hypotheses, sharing data, and sharing other analytical material. Recent research found that when journals begin offering badges, the rate at which the manuscripts they publish rely on these open science practices increases substantially (Kidwell et al. 2016). Finally, journals might publish preregistered replications. Prior to collecting data and doing analyses, authors submit to the journal a study justification and replication protocol for review. If all looks well, a tentative acceptance may be granted. After collecting data, doing analyses, and writing up results to describe whether the original findings were replicated, the full manuscript is resubmitted to the journal for a cursory review, preferably by the initial reviewers. Overall, articles are the primary means of disseminating research results, which makes journals the main depository of evidence. Journals are therefore responsible for helping correct errors in the scientific record due to publication bias and must play a crucial role in encouraging a culture of replication.

Proreplication Efforts in Criminology

Criminology has not directly and systemically addressed replication and the move to an open science framework. Closely related fields like psychology, political science, economics, and sociology have initiated at least some concrete steps toward these goals. We believe now is the time for criminology, and we mention here a few proreplication efforts in the field and in its agencies and journals that suggest criminologists are receptive.

[4]However, the research community is equally responsible for fetishizing an original finding, which seems to hold a talismanic effect and take on mystical unassailable status simply because it was first. Careful follow-up studies that find no association and replications that do not replicate original findings are often ignored while the original study continues to be cited. The clear existence of publication bias and the decline effect make this unreasonable.

The Campbell Collaboration, which originated in 2000, is patterned on the Cochrane Collaboration. Its aim is to "promote positive social and economic change through the production and use of systematic reviews and other evidence synthesis for evidence-based policy and practice" (**https://www.campbellcollaboration.org/about-campbell/vision-mission-and-principle.html**). This organization is external to criminology, and meta-analysis is not a substitute for replication (van Elk et al. 2015). Still, many Campbell reviews cover crime and justice topics, many criminologists have prepared reviews for them, and criminology's interaction with the Campbell Collaboration signals recognition of the need for more systematic assessment of our evidence base. At the disciplinary level, the recently adopted *Code of Ethics* of the American Society of Criminology (2016) contains a statement about sharing data and workflow for replication purposes. The American Society of Criminology and the Academy of Criminal Justice Sciences might also consider hosting conference workshops and panels on undertaking replications and reproducible research. More generally, because there is no wholly agreed upon definition of reproducibility or replication, a more formal discussion of what form these practices should take in criminology is required.

The National Institute of Justice (NIJ) hosts the website crimesolutions.gov, which is an agency-level example of efforts "to inform practitioners and policy makers about what works, what doesn't, and what's promising in criminal justice, juvenile justice, and crime victim services" (**https://www.crimesolutions.gov/about.aspx**). Although the information on the site is not usually based on replications, this service reflects the need for greater and more careful assessment of programs and research quality and for disseminating that information broadly. Given this commitment, agencies like NIJ and others should also devote more funding to replication research and consider requiring funded projects to carry out reproducible research when possible. If this is done, these agencies will have greater confidence in the information they share with policy makers and practitioners.

Criminology journals must be leaders in cultivating a culture of replication in the field. That this inaugural volume of the *Annual Review of Criminology* contains an article on replication is important. This year, an issue of the *Journal of Contemporary Criminal Justice* will also be devoted to replication work in criminology and criminal justice. These are encouraging signs, but our research presented above (see also McNeely & Warner 2015) showed a miniscule proportion of articles in criminology journals are replications. Our journals must address publication bias and publish more replications and more null findings. Our leading journals might also consider proreplication policies on data and workflow like those of leading journals in allied fields such as political science (e.g., *American Journal of Political Science*), psychology (e.g., *Psychological Science*), and economics (e.g., *American Economic Review*).

CONCLUSION

Crime and justice are part of everyday social life and are often highly politicized. Further, much research in criminology is translatable to policy and practice, which can be expensive and have unintended harmful consequences. Thus, the trustworthiness of criminology's scientific record is not only fundamental to its sustainability as a field but also integral to its relevance and reputation. Replication plays an essential self-correcting role in the systematic accumulation of knowledge by ensuring original findings are not due to error or chance. It also provides a deterrent effect: If one must share data and workflow, and if there is reasonable certainty one's research will be replicated by others, then QRPs will be less likely. To date, however, replication has been very rare in criminology, and publication bias has likely skewed the evidence base. When other fields faced up to this combination of limitations they discovered reproducibility problems. There are large literatures that provide detailed and nuanced treatments of replication and that describe

how other disciplines are addressing these setbacks. We hope this review provides a worthwhile introduction to replication for criminology and other social sciences.

To paraphrase a classic statement on the subject in psychology (Cohen 1994, p. 997), for generalization criminologists must rely on replication. Criminology should encourage a culture that supports open science, funds more replications, carries out and publishes more replications, and advocates for policy and practice based on reproducible findings. Although replication is only one of many ways to a reliable scientific record, criminology will be stronger and held in higher esteem the more the field, professional organizations, funders, journals, and individual researchers adopt these practices.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Rachel Levin for research assistance on this project.

LITERATURE CITED

Am. Soc. Criminol. 2016. *Code of Ethics*. Columbus, OH: Am. Soc. Criminol. **http://asc41.com/code_of_ethics_copies/ASC_Code_of_Ethics.pdf**

Anderson CJ, Bahník S, Barnett-Cowan M, Zuni K. 2016. Response to comment on "Estimating the reproducibility of psychological science." *Science* 351:1037

Baker M. 2016. Is there a reproducibility crisis? *Nature* 533:453–54

Baumeister RF. 2002. Ego-depletion and self-control failure: an energy model of the self's executive function. *Self Identity* 1:129–36

Baumeister RF, Bratslavsky E, Muraven M, Tice DM. 1998. Ego depletion: Is the active self a limited resource? *J. Personal. Soc. Psychol.* 74:1252–65

Begley CG, Ellis LM. 2012. Raise standards for preclinical cancer research. *Nature* 483:531–33

Bem D. 2011. Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *J. Personal. Soc. Psychol.* 100:407–25

Berk RA, Campbell A, Klap R, Western B. 1992. A Bayesian analysis of the Colorado Springs spouse abuse experiment. *J. Crim. Law Criminol.* 83:170–200

Brandt MJ, IJzerman H, Dijksterhuis A, Farach F, Geller J, et al. 2014. The replication recipe: What makes for a convincing replication? *J. Exp. Soc. Psychol.* 50:217–24

Brodeur A, Lé M, Sangnier M, Zylberberg Y. 2016. Star wars: The empirics strike back. *Am. Econ. J. Appl. Econ.* 8:1–32

Burger JM. 2009. Replicating Milgram: Would people still obey today? *Am. Psychol.* 64:1–11

Camerer CF, Dreber A, Forsell E, Wu H. 2016. Evaluating replicability of laboratory experiments in economics. *Science* 351:1433–36

Chavalarias D, Wallach JD, Li AH, Ioannidis JP. 2016. Evolution of reporting P values in the biomedical literature. *JAMA* 315:1141–48

Christakis DA, Zimmerman FJ. 2013. Rethinking reanalysis. *JAMA* 310:2499–500

Cohen J. 1994. The earth is round ($p < .05$). *Am. Psychol.* 49:997–103

Collins HM. 1985. *Changing Order: Replication and Induction in Scientific Practice*. London: Sage

Condon DM, Graham E, Mroczek DE. 2017. Replication research. In *Wiley-Blackwell Encyclopedia of Personality and Individual Differences*, ed. BJ Carducci. Malden, MA: Wiley-Blackwell. In press

de Bono S. 2005. Is the spirit of Piltdown Man alive and well? *The Telegraph*, Sept. 7, **http://www.telegraph.co.uk/technology/3342867/Is-the-spirit-of-Piltdown-man-alive-and-well.html**

Dolínski D, Grzyb T, Folwarczny M, Trojanowski J. 2017. Would you deliver an electric shock in 2015? Obedience in the experimental paradigm developed by Stanley Milgram in the 50 years following the original studies. *Soc. Psychol. Personal. Sci.* In press. **http://journals.sagepub.com/doi/10.1177/1948550617693060**

Donn J. 2016. Medical benefits of dental floss unproven. *Associated Press*, Aug. 2, **http://bigstory.ap.org/article/f7e66079d9ba4b4985d7af350619a9e3/medical-benefits-dental-floss-unproven**

Fanelli D. 2010. "Positive" results increase down the hierarchy of the sciences. *PLOS ONE* 5(4):e10068

Fanelli D. 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90:891–904

Feynman R, Leighton R. 1985. *Surely You're Joking Mr. Feynman! Adventures of a Curious Character*. New York: Norton

Fleischmann M, Pons S. 1989. Electrochemically induced nuclear fusion of deuterium. *J. Electroanal. Chem.* 261:301–8

Franco A, Malhotra N, Simonovits G. 2014. Publication bias in the social sciences: unlocking the file drawer. *Science* 345:1502–5

Freedman LP, Cockburn IM, Simcoe TS. 2015. The economics of reproducibility in preclinical research. *PLOS Biol.* 13(6):e1002165

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

Gilbert DT, King G, Pettigrew S, Wilson TD. 2016. Comment on "Estimating the reproducibility of psychological science." *Science* 351:1037

Grahe JE, Reifman A, Hermann AD, Wiebe RP. 2012. Harnessing the undiscovered resource of student research projects. *Perspect. Psychol. Sci.* 7:605–7

Guallar E, Stranges S, Mulrow C, Appel LJ, Miller ER III. 2013. Enough is enough: Stop wasting money on vitamin and mineral supplements. *Ann. Intern. Med.*159:850–51

Hagger MS, Chatzisarantis NLD, Alberts H, Zwienenberg M. 2016. A multilab preregistered replication of the ego-depletion effect. *Perspect. Psychol. Sci.* 11:546–73

Hannon L, Defina R. 2016. Can incarceration really strip people of racial privilege? *Sociol. Sci.* 3:190–201

Herndon T, Ash M, Pollin R. 2014. Does high public debt consistently stifle economic growth? A critique of Reinhart & Rogoff. *Camb. J. Econ.* 38:257–79

Hureau DM. 2016. *Essays exploring urban violence*. PhD Diss., Harvard Univ., Cambridge, MA

Hureau DM, Braga AA. 2018. The trade in tools: the market for illicit guns in high-risk networks. *Criminology*. In press

Ioannidis JPA. 2005a. Why most published research findings are false. *PLOS Med.* 2(8):e124

Ioannidis JPA. 2005b. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 294:218–28

Ioannidis JPA. 2008. Why most discovered true associations are inflated. *Epidemiology* 19:640–48

Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. 2016. Reproducible research practices and transparency across the biomedical literature. *PLOS Biol.* 14(1):e1002333

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

Kahneman D. 2014. A new etiquette for replication. *Soc. Psychol.* 45:310–11

Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, et al. 2016. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLOS Biol.* 14(5):e1002456

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

Kline RB. 2004. *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. Washington, DC: Am. Psychol. Assoc.

Kunert R. 2016. Internal conceptual replications do not increase independent replication success. *Psychon. Bull. Rev.* 23:1631–38

Lall R. 2016. How multiple imputation makes a difference. *Political Anal.* 24:414–33

Lauritsen JL. 2005. Explaining patterns of offending across the life course: Comments on interactional theory and recent tests based on the RYDS-RIS data. *Ann. Am. Acad. Political Soc. Sci.* 602:212–28

LeBel EP. 2015. A new replication norm for psychology. *Collabra* 1:4. **https://www.collabra.org/articles/10.1525/collabra.23/**
Leek JT, Peng RD. 2015. Reproducible research can still be wrong: adopting a prevention approach. *PNAS* 112:1645–46
Lemons CJ, King SA, Davidson KA, Berryessa TL, Gajjar SA, Sacks LH. 2016. An inadvertent concurrent replication. Same roadmap, different journey. *Remedial Spec. Educ.* 37:213–22
Lieberson S, Lynn FB. 2002. Barking up the wrong branch: scientific alternatives to the current model of sociological science. *Annu. Rev. Sociol.* 28:1–19
Long JS. 2009. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press
Makel MC, Plucker JA, eds. 2017. *Toward a More Perfect Psychology: Improving Trust, Accuracy, and Transparency in Research*. Washington, DC: Am. Psychol. Assoc
Makel MC, Plucker JA. 2014a. Facts are more important than novelty: replication in the education sciences. *Educ. Res.* 43:304–316
Makel MC, Plucker JA. 2014b. Creativity is more important than novelty: reconsidering replication as a creativity act. *Psychol. Aesthet. Creativity Arts* 8:27–29
Makel MC, Plucker JA, Freeman J, Lombardi A, Simonsen B, Coyne M. 2016. Replication of special education research: necessary but far too rare. *Remedial Spec. Educ.* 37:205–12
Makel MC, Plucker JA, Hegarty B. 2012. Replications in psychology research: How often do they really occur? *Perspect. Psychol. Sci.* 7:537–542
Maniadis Z, Tufano F, List JA. 2014. One swallow doesn't make a summer: new evidence on anchoring effects. *Am. Econ. Rev.* 104:277–90
McBee M, Makel M, Peters SJ, Matthews MS. 2017. *A Manifesto for Open Science in Giftedness Research*. Charlottesville, VA: Cent. Open Sci. **https://osf.io/qhwg3/**
McNeely S, Warner JJ. 2015. Replication in criminology: necessary practice. *Eur. J. Criminol.* 12:581–97
Milgram S. 1963. Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67:371–78
Munafo MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, et al. 2017. A manifesto for reproducible science. *Nat. Hum. Behav.* 1:0021
Neth. Organ. Sci. Res. 2017. *Repeating Important Research Thanks to Replication Studies*. The Hague, Neth.: NWO. **https://www.nwo.nl/en/news-and-events/news/2017/social-sciences/repeating-important-research-thanks-to-replication-studies.html**
Nichols T. 2017. *The Death of Expertise: The Campaign Against Established Knowledge and Why it Matters*. New York: Oxford Univ. Press
Nosek BA, Alter G, Banks GC, Yarkoni T. 2015. Promoting an open research culture. *Science* 348:1422–25
Nuijten MB, Hartgerink CH, van Assen MA, Epskamp S, Wicherts JM. 2016. The prevalence of statistical reporting errors in psychology (1985–2013). *Behav. Res. Methods* 48:1205–26
Nuzzo R. 2015. How scientists fool themselves—and how they can stop. *Nature* 526:182–85
O'Boyle EH, Banks GC, Gonzalez-Mulé E. 2014. The chrysalis effect: how ugly initial results metamorphosize into beautiful articles. *J. Manag.* 43:376–99
Open Sci. Collab. 2015. Estimating the reproducibility of psychological science. *Science* 349:aac4716
Open Sci. Framew. 2017. *Course Syllabi for Open and Reproducible Methods*. Charlottesville, VA: Cent. Open Sci. **https://osf.io/vkhbt/**
Patil PP, Peng RD, Leek JT. 2016. What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspect. Psychol. Sci.* 11:539–44
Petrasso RD, Chen X, Wenzel KW, Parker RR, Li CK, Fiore C. 1989. Problems with the γ-ray spectrum in the Fleischmann et al. experiments. *Nature* 339:183–85
Pickett JT, Roche SP. 2017. Questionable, objectionable or criminal? Public opinion on data fraud and selective reporting in science. *Sci. Eng. Ethics.* **https://doi.org/10.1007/s11948-017-9886-2**
Pigott TD, Valentine JC, Polanin JR, Williams RT, Canada DD. 2013. Outcome-reporting bias in education research. *Educ. Res.* 42:424–32
Pridemore WA. 2011. Poverty matters: a reassessment of the inequality-homicide relationship in cross-national studies. *Br. J. Criminol.* 51:739–72
Pridemore WA. 2008. A methodological addition to the cross-national empirical literature on social structure and homicide: a first test of the poverty-homicide thesis. *Criminology* 46:133–54

Prinz F, Schlange T, Asadullah K. 2011. Believe it or not: How much can we rely on published data on potential drug targest? *Nat. Rev. Drug Discov.* 10:712–13

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

Sampson RJ, Winship C, Knight C. 2013. Translating causal claims: principles and strategies for policy-relevant criminology. *Criminol. Public Policy* 12(4):587–616

Saperstein A, Penner AM. 2010. The race of a criminal record: how incarceration colors racial perceptions. *Soc. Probl.* 57:92–113

Schmidt S. 2009. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev. Gen. Psychol.* 13:90–100

Schmidt S. 2017. Replication. In *Toward a More Perfect Psychology: Improving Trust, Accuracy, and Transparency in Research*, ed. M Makel, JA Plucker, pp. 233–54. Washington, DC: Am. Psychol. Assoc.

Schwarz N, Clore GL. 2016. Evaluating psychological research requires more than attention to the *N*: a comment on Simonsohn's (2015) "small telescopes." *Psychol. Sci.* 27:1407–9

Shapin S, Schaffer S. 1985. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton Univ. Press

Sherman LW, Berk RA. 1984. The specific deterrent effects of arrest for domestic assault. *Am. Sociol. Rev.* 49:261–72

Sherman LW, Schmidt JD, Rogan DP, Gartin PR, Cohn EG, et al. 1991. From initial deterrence to longterm escalation: short-custody arrest for poverty ghetto domestic violence. *Criminology* 29:821–50

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

Simmons JP, Nelson LD, Simonsohn U. 2012. A 21 word solution. **http://dx.doi.org/10.2139/ssrn.2160588**

Simons DJ, Shoda Y, Lindsay DS. 2017. *Constraints on Generality (COG): A Proposed Addition to All Empirical Papers*. Charlottesville, VA: Cent. Open Sci. **https://osf.io/phptw/**

Simonsohn U. 2015. Small telescopes: detectability and the evaluation of replication results. *Psychol. Sci.* 26:559–69

Simonsohn U. 2016. Each reader decides if a replication counts: reply to Schwarz and Clore. *Psychol. Sci.* 27:1410–12

Spellman B, Gilbert EA, Corker KS. 2017. *Open Science: What, Why, and How*. Charlottesville, VA: Cent. Open Sci. **https://osf.io/8u3qe/**

Srivastava S. 2012. A pottery barn rule for scientific journals. *The Hardest Science*, Sept. 27, **https://hardsci.wordpress.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/**

Stodden V, McNutt M, Bailey DH, Dellman E, Gil Y, et al. 2016. Enhancing reproducibility for computational methods. *Science* 354:1240–41

Thornberry TP. 2005. Notes on theory construction and theory testing: a response to Osgood and Lauritsen. *Ann. Am. Acad. Political Soc. Sci.* 602:229–39

Tittle C. 2004. The arrogance of public sociology. *Soc. Forces* 82:1639–43

van Elk M, Matzke D, Gronau QF, Guan M, Vanderkerckhove J, Wagenmakers E-J. 2015. Meta-analyses are no substitute for registered replications: a skeptical perspective on religious priming. *Front. Psychol.* 6:1365

Van Noorden R. 2011. The trouble with retractions. *Nature* 478:26–28

Wellford CF. 2010. Criminologists should stop whining about their impact on policy and practice. In *Contemporary Issues in Criminal Justice Policy: Policy Proposals from the American Society of Criminology Conference*, ed. NA Frost, JD Freilich, TR Clear, pp. 17–23. Belmont, CA: Wadsworth Cengage Learn.

Wicherts JM, Borsboom D, Kats J, Molenaar D. 2006. The poor availability of psychological research data for reanalysis. *Am. Psychol.* 61:726–28

Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM. 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. *Front. Psychol.* 7:1832

Young C, Horvath A. 2015. Sociologists need to be better at replication. *orgtheory.net*, Aug. 11, **https://orgtheory.wordpress.com/2015/08/11/sociologists-need-to-be-better-at-replication-a-guest-post-by-cristobal-young/**