Doing Better Science:
Improving Review & Publication Protocols to Enhance the Quality of Criminological Evidence

"Our job as scientists is to discover truths about the world…. Although we aspire to be accurate, errors are inevitable"
(Simmons et al. 2011, p.1359).

This is not our grandfather's science. Extraordinary technological innovations, including the Internet and World Wide Web, increased computer storage and computational capacities, have increased speed and sophistication of research. Perhaps most striking is the effect of these innovations on the exponential increase in the quantity of publications, which continues to grow (~5% per year) and shows no sign of slowing (Sarewitz 2016). The sheer quantity of published research is astounding. More than 4,000 new criminology articles are now published per year (Koehler 2020). There are more of us, doing more research, faster than ever before. And, still, we face pressure to do more (Sarewitz 2016). This pressure to publish has not only intensified, it has expanded beyond faculty at research universities to faculty in schools without graduate programs, to graduate students for procuring jobs, and even to undergraduates for graduate school admissions (Nosek et al. 2012).

Although technological innovations have transformed scientific research, key facets of science remain firmly 'rooted in anachronistic practices of bygone eras' (Nosek & Bar-Anan 2012, p.217). Publications remain the primary way of communicating scientific ideas and empirical findings, and they remain the currency of scientific prestige and the means to job security (Nosek & Lakens 2014). 'Classical peer review'—a non-transparent, iterative process which is inefficient and improvable (discussed here)—continues to be the way most papers are evaluated for publication (Walker & da Silva 2015). Not all papers are considered equally worthy of publication, in general, or for increasingly competitive pages in high-prestige journals in particular, and not all papers get published (Nosek et a. 2012).[1] Publication bias is pervasive. Studies with novel and positive (significant) findings and which are framed in a tidy package are favored for publication over null findings and replications, especially in high-impact journals where rejection rates can exceed 90% (Fanelli 2012).

To succeed in this increasingly competitive milieu, researchers are pressured to publish more studies in the highest impact journals as possible, with professional incentives—prestige, salary, funding—reflecting this prioritization (Sarewitz 2016). These pressures create a problematic imbalance, akin to Durkheim's notion of social anomie, where the emphasis on the ends (publishing more studies) is out of balance with emphasis on the means (rigorous research practices), leading to a definite pressure on researchers to deviate by employing 'questionable research practices' (QRPs)as well as in more blatant ways through fabrication, falsification, and plagiarism (FFP) (Stenek 2006). The incentives for 'getting it right' are incommensurate with the incentives for getting it published (Nosek et al. 2012), and 'the result is a kind of [scientific] arms race in which researchers who play strictly by the rules are at a competitive disadvantage' (John et al. 2012, p.524).

Although widely publicized examples of FFPs are obviously problematic, evidence suggests that less egregious but more common QRPs may have a more significant role in distorting the literature (John et al. 2012). QRPs include practices such as HARKing (hypothesizing after results are known; Kerr 1998) and '*p*-hacking', (also known as data dredging, data mining, snooping, fishing, significance chasing, and double dipping), which means trying multiple things until you get the desired *p*-value results and then creating a story post hoc (Simmons et al. 2011).[2] These practices, in which

---

[1] Furthermore, given the copious number of studies, limited time and information, the primary heuristic for evaluating the quality of scientific findings, as least in the short run, remains the prestige of the journal in which papers are published (Kriegeskorte 2012). Given the drastic increase in the number of studies relatively stable numbers of 'high prestige' publications, the competition for increasingly scarce space in high-impact journals is escalating.

[2] Although a thorough discussion of the reasons for these practices is out of scope, they include inadequate researcher training in scientific methods, the aforementioned unbalanced incentives, and a scientific culture that rewards, or at least tacitly allows, non-transparency.

researchers employ hidden 'researcher degrees of freedom' and multiple testing, significantly increase Type 1 errors and inflate the effects sizes of genuine effects (Simmons et al. 2011). Common tools of statistical inference designed for standard confirmatory null hypothesis are not diagnostic in the presence of p-hacking and provide false comfort about the reliability and credibility of findings (Nosek et al. 2018). In some cases, the employment of QRPs makes a false-positive finding more likely than not (Simmons et al. 2011). To remedy this situation, "[w]e need measures to counter the natural tendency of enthusiastic scientists who are motivated by discovery and the pressure to publish more too see or create patterns out of noise" (Munafo et al. 2017).

Conflict of Interest Problem: Not New but Getting Worse

This situation—where even more scholars are expected to produce more publications with innovative and significant findings—involves a conflict of interest between the aims of science (the production of robust knowledge) and the careers of particular scientist (the publication of more papers in high-impact journals). This conflict of interest is not new, but the current situation is unique and pressing (Fanelli 2012). Expanding pressures for publication are accompanied by increased competition for scarce pages in high prestige journals and a worsening Ph.D. supply—job-demand ratio. Moreover, advances in computer technology, which facilitate rapid and automated data analysis (facilitating multiple testing with a few clicks of a button), allow researchers to quickly and easily engage in QRPs that depart from standard null hypothesis testing and make reported p-values non-diagnostic. This increased ease of multiple testing combined with our apophenia (our tendency to create patterns out of noise), confirmation and hindsight biases greatly escalate the risk of false positives, even in the absence of ill intent (Munafo et al. 2017). If we hold constant the proportion of false (spurious, chance, inaccurate, even fraudulent) findings, and there are reasons to believe that the proportion of false positives are increasing (Fanelli 2012), the exponential increase in the quantity of published research means we also have an exponential increase in the publication of false-positive findings. As a consequence, as Ioannidis (2005) famously noted, most published findings are probably false positives (Simmons et al., 2018).

This matters for science, for criminology, and for society. Publication bias (against non-significant findings) and the publication of false-positive findings, whether by honest mistake or fraudulent activity, distorts the literature by producing a glut of false positive findings and an absence of negative findings. This impedes the accumulation of knowledge, wastes resources, and undermines public trust in scientific research (and thus evidence-based policies) (e.g., Nosek et al. 2012). Quite often, novel, positive findings do not go quietly into the night even after multiple studies fail to replicate them, even in the presence of research errors and fabrication (John et al. 2012; Simmons et al. 2011).

Following a spate of replicability crises, several high-profile retractions (e.g., Wakefield MMR study in the *Lancet*), and a surfeit of conflicting results that seem to support any position one wishes to take ('backed by science'), recent years have seen the erosion of trust in scientific evidence, by both scientists and the general public (e.g., NAS 2017). More people reject scientific evidence as untrustworthy than in the past. Science is crucially dependent on public support, which is rooted not in a belief that science is infallible but in the ability of the scientific process to correct mistakes in a manner that gets us closer to the truth (Kerr 1998).

In my view and that of others, this state of affairs is unacceptable and unsustainable. We must address these issues to improve the quality[3] of published research in criminology. There are several fundamental facets of modern science and academia that should be significantly reworked to foster better science, perhaps most fundamentally the structure of incentives (quality over quantity; collaboration versus competition) and the administration of funding (see Sarewitz 2016, also Nosek et al. 2012). Here, I focus on one specific leverage point to illustrate current limitations and suggest

---

[3] Higher quality research, as I use it here, means research that is robust, replicable, not dependent on arbitrary researcher decisions or omitted variable bias, and has a clear connection to extant knowledge and, in some way, has potential to advance knowledge, which includes both novel findings and replications, which increase confidence in the truth value of a claim (see Pridemore et al. 2018).

alternatives in order to stimulate discussion: peer-review and publication processes. This is a site of relative stasis in criminology, ripe for change.

Although concerns about peer review have long been expressed, over the past decade, scholars have suggested and initiated strategies to improve peer review in ways that have several benefits including markedly reducing, if never completely eliminating, the publication of false findings and publication biases (e.g., Munafo et al. 2017; Nosek et al. 2012). Several journals are already implementing these strategies (see cos.io). Drawing on these ideas, I propose for consideration a series of standards intended to improve peer review and promote transparency.

Classical Peer Review: Largely Unchanged and Not Up for the Contemporary Job

The conflict of interest problem is compounded by the inadequacy of classical peer review (CPV) for catching errors or QRPs in research. The system of CPV (blind review by 2-6 reviewers who make recommendations to editors who make decisions) was designed to work well in a system where scientists work (unselfishly) in pursuit of robust, replicable knowledge—serving as their 'own toughest critics'. However, this is not our world. In our world, where scientists face pressures to 'publish or perish', the quest for truth can take a backseat to personal interests in the service of employment or job stability.

Peer review is notoriously unreliable and error-prone. Research has demonstrated the failures of CPV to detect major methodological errors and a lack of agreement among reviewers that is only slightly better than chance. The inadequacy of peer review is a largely function of two (changeable) factors: lack of incentives for high quality review and insufficient information. Focusing on the former, the increase in research quantity, means more papers need reviewing, which increases the time constraints on reviewers. This, along with the general lack of incentives for reviewing, and high-quality reviews, in particular, combine to make rigorous, time-consuming reviews a disincentivized (even curiously unselfish) act for the individual scholar. Thorough reviews of 10k+ word papers with sophisticated methods cannot (usually) be conducted in a few hours' time but require a more sustained time and intellectual commitment. However, time spent on reviews is time not spent on one's own research and teaching. Consequently, peer review is often a rushed activity that does not provide the evaluation or gatekeeping that we expect it to (and think it does).

In addition to lack of incentives, there is a lack of information. Referees can only evaluate a study based on the information that the authors provide. *Criminology*, like other disciplinary journals, does not provide standards for disclosure. Consequently, information on methods is usually incomplete and post-hoc decisions (e.g., how dealt with outliers, how many tests conducted) are invariably not fully described. Furthermore, even in the presence of full disclosure, reviewers cannot ascertain whether mistakes were made, including basic errors in coding, given that only summary information is provided. Research errors can occur in data cleaning, coding, data analysis, and reporting, and almost none of these can be detected only with the summary report of the findings in a manuscript (Nosek et al. 2012). Although the actual rate of errors in scientific research is unknown, evidence suggests that it is unacceptably high (Bakker & Wicherts 2011).

Science is distinguished from other ways of knowing based on the ability to verify claims and disprove them with empirical evidence. Right now, criminology is not operating in a manner that facilitates verification and replication. But it does not have to be this way.

Motivating Change by Requiring Change

Below I submit for consideration a series of protocols around review and publication processes aimed at improving the transparency, robustness and replicability of scientific research. These are but a few suggestions to stimulate a discussion, most of which have been suggested by others (see, e.g., Munafo et al. 2017; Simmons et al. 2011), and some of which are already employed by journals with beneficial effect. Notably, employing these protocols will require many

of us to revise our research processes, and some will slow us down (at least at first). However, with 4000+ papers published per year in criminology, we do not need to rush.

The following protocols and policies are proposed as *requirements* for publication in, at a minimum, ASC journals. Collectively, we have demonstrated that strong encouragement is not enough; mandates are necessary to gain adherence.

Author Requirements (*Aims: increase detection of errors and accountability, and reduce the incentives for QRPs and FFPs because more easily detected*):

(1) **Pre-registration of analysis plans for all confirmatory research**. In pre-registration, researchers detail their hypotheses, methods, and planned analyses before the initiation of their research. Effective preregistration of study or analysis protocols involves specific, detailed plans, and how decisions will be made that are verifiable and thus minimize QRPs (given they will be detectable). In manuscript submission forms, authors will indicate whether the paper is confirmatory or exploratory. If confirmatory, time-stamped study preregistration will be a requirement of submission (see Nosek et al. 2018, also Munafo et al. 2017.)

(2) **Disclosure of all methodological decisions and author confirmation of full disclosure** (e.g., reporting all of the inclusion and exclusion criteria, variables, manipulations, tests, etc.; see Simmons et al. 2011; 2018). Author checklists are increasingly common in various journals (e.g., the MDAR open science checklist for life sciences), and could be easily refined for criminology.

(3) **Provision of data and code to reviewers *at the point of review***. In submission forms, authors will indicate whether articles are empirical or non-empirical; if empirical, the data and code are required to be uploaded. Submissions without data and code *are not accepted* for review.

(4) **Publication of data and code alongside article**. At the time of publication, code and data (which was made available at review, along with all data and code used in revisions) will be published alongside articles on the publishers website, *available to all*. This is already done in some journals, where data is provided as a supplement. In instances where data is deemed too sensitive for public sharing and or is too large, authors can deposit data in one of several established repositories, accessible to bona fide researchers upon review of their credentials. This 'sensitive exception' must not be used to skirt open data policies; data must be shown to be available *prior* to the acceptance of a manuscript for publication.[4]

(5) **Inclusion of author contribution sections**, which clarify each authors contribution to a specific manuscript, including specific responsibilities. This is also increasingly common and informative. This may be of particular benefit to junior scholars with coauthors whose contributions are limited to provision of data.

Review/Publisher Protocols (*Aims: increase transparency, confirm analysis results, incentivize higher quality reviews and provide credit for those who provide such reviews*)

(6) **Creation of a new 'analysis corroborator' reviewer role**. This reviewer is tasked with reproducing the study findings with the data and code provided by the authors to assess their veracity, to scrutinize any arbitrary coding and/or analysis decisions, and catch any mistakes. This reviewer who conducts the validation analyses

---

[4] Considering how to make qualitative data available is not as straightforward given some of the unique facets of qualitative data analysis. At a minimum, all coding or researcher decisions should be made public, such that the findings could be fully reproduced *if* the data were available. And scholars must commit to sharing such data with bona fide researchers under certain circumstances (providing appropriate protections, including commitment to maintaining the privacy and confidentiality of all sensitive information).

should be given credit for their effort; their name and role as 'analysis corroborator' should be listed on the publication, and this effort should count as a significant research contribution.

(7) **Provision of time-stamped paper histories, initial submissions, reviews, editor comments, and author responses to comments**. This provides transparency and also maintains a reviewable record and teaching materials. Importantly, referee reviews can be *signed or anonymous* and are citable contributions to the literature, created with their own DOIs. In this form, the reviews are blinded, but reviewers can choose to reveal their identities after the review process to gain credit for their often useful contributions and scholarly effort.

As discussed above, quality reviews are time consuming and mostly invisible contributions to science. Despite this, many scholars provide insightful, careful reviews whose impacts are, unfortunately, limited given that they are for the eyes of the authors, editors, and other reviewers only (Nosek & Bar-Anan 2012). To provide transparency, incentivize quality reviews and provide credit (as a scholarly contribution) and reputational credit for scientists who provide consistently high-quality reviews, we should publish these reviews. Additionally, published reviews can provide useful information about the identified strengths and limitations of the paper, and serve as training materials for other referees (Bravo et al. 2019). Notably, various forms of OPR are already practice in several journals, in full or in part, and OPR is gaining in support (Walker & da Silva 2015). Perhaps surprisingly, there is no evidence that anonymous review is superior to eponymous review, and several studies over the past decade suggest that OPR does not undermine the quality of the review (and may slightly enhance it) (e.g., Bravo et al. 2019).

Conclusion

"We should embrace disclosure and preregistration as if the credibility of our profession depended on it. Because it does."
(Simmons et al. 2018, p.259)

The goal of this essay is to stimulate a discussion around reforming a system that is not operating as well as it should. Although in my view, we must reform or adjust the system, some may argue that we do not need to do things differently, arguing either that criminology does not have a problem with the publication of false-positive findings and publication bias or that that we have always had these problems, and it isn't serious. But we do have a problem; evidence suggests it is serious; and we already are doing some things differently. Our discussion should focus on how to improve things not whether we should. There are a variety of innovations, and we now have the technical means to implement them. If we want criminological research to enhance our understanding of the world around us, we must make it easier for scientists to publish true facts than exciting, significant, and ultimately false results (Simmons et al. 2018). Of course, we can continue to do things as we are, but to whose benefit and at what cost?

References

Bakker, Marjan, & Jelte M. Wicherts. 2011. The (mis) reporting of statistical results in psychology journals. *Behavior Research Methods* 43(3): 666-678.

Bravo, Giangiacomo, Francisco Grimaldo, Emilia López-Iñesta, Bahar Mehmani, & Flaminio Squazzoni. 2019. The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications* 10(1): 1-8.

Fanelli, Daniele. 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90(3): 891-904.

Ioannidis, John PA. 2005. Why most published research findings are false. *PLos Med* 2(8): e124.

John, Leslie K., George Loewenstein, & Drazen Prelec. 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science* 23: 524-532.

Kerr, Norbert L. 1998. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review* 2(3): 196-217.

Koehler, Johann. 2020. Contours of Criminological Thought. *Unpublished manuscript.*

Kriegeskorte, Nikolaus. 2012. Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. *Frontiers in Computational Neuroscience* 6: 79.

Munafò, Marcus R., *et al.*. 2017. A manifesto for reproducible science. *Nature Human Behaviour* 1(1): 1-9.

National Academies of Sciences, Engineering, and Medicine. 2017. Examining the Mistrust of Science: Proceedings of a Workshop—in Brief. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/28650594.

Nosek, Brian A., & Yoav Bar-Anan. 2012. Scientific utopia: I. Opening scientific communication. *Psychological Inquiry* 23(3): 217-243

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, & David T. Mellor. 2018. The preregistration revolution. *PNAS* 115: 2600-6.

Nosek, Brian A., & Daniël Lakens. 2014. A Method to Increase the Credibility of Published Results. *Social Psychology* 45(3): 137-141.

Nosek, Brian A., Jeffery R. Spies, & Matt Motyl. 2012. Scientific Utopia II: Restructuring Incentives and Practices to Promote Truth over Publishability. *Perspectives on Psychological Science* 7(6): 615-631.

Pridemore, William Alex, Matthew C. Makel, & Jonathan A. Plucker. 2018. Replication in criminology and the social sciences. *Annual Review of Criminology* 1: 19-38.

Sarewitz, Daniel. 2016. The pressure to publish pushes down quality. *Nature*. 533: 147.

Simmons, Joseph P., Leif D. Nelson, & Uri Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science* 22: 1359-1366.

Simmons, Joseph P., Leif D. Nelson, & Uri Simonsohn. 2018. False-positive citations. *Perspectives on Psychological Science* 13(2): 255-9.

Steneck, Nicholas H. 2006. Fostering Integrity in Research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics* 12(1): 53-74.

Walker, Richard, & Pascal Rocha da Silva. 2015. Emerging trends in peer review—a survey. *Frontiers in Neuroscience* 9: 169.