When You Hit a Fork in the Road, Take It: What the Latest Controversies and Data Tell Us About Our Field, Open Science, and the Way Forward

*John Paul Wright, Ph.D.**
School of Criminal Justice
University of Cincinnati

Abstract

The latest controversies in our field highlight the chasm between the pursuit of rigorous, replicable, science and disciplinary incentives that reward the publication of novel findings, the use of questionable research strategies, and a preference for ideologically uniform narratives. To advance our discipline, I argue we should embrace the open science movement and leverage the empirical work that highlights the limits of social science generally, and criminology specifically. Embracing the open science movement, however, is more than changing the mechanics of our science. It will also require a cultural change -- one that prioritizes the pursuit of truth over all else.

---

* This article is based on professor Wright's keynote lecture delivered at the International Forum of the Korean Institute of Criminology, Seoul Korea, 6 December, 2019.
* http://dx.doi.org/10.36889/IJCJ.2020.12.2.2.001
* Received 6 December 2019; Received in revised form 20 October 2020; Accepted 5 November 2020

INTRODUCTION

Criminology is perhaps better situated today, more than ever before, to provide answers to thorny policy issues and to make meaningful contributions to science on etiology. We have more Ph.D. programs and thus people trained in the discipline, more datasets from which to draw on and to analyze, more statistical techniques to dazzle readers with, more journals to publish in and consequently, more studies to read and to dissect. Our discipline, once shunned by other fields and disparaged as nothing more than a marginal offshoot of sociology, has by any measure, captured widespread intellectual attention and intellectual legitimacy. Today, criminologists from around the world contribute to an ongoing dialogue about crime, criminality, and the control of wayward behavior. Criminology, it seems, is at the height of its glory and influence.

Given our gains, it may seem odd to sound a warning about our future but that is precisely what I'm going to do. As many businesses have learned, often through insolvency, growth is relatively easy compared to maintaining a competitive edge or expanding further market shares. Examples are all around us of organizations, even academic disciplines, moving leaps and bounds ahead of others only to reach an apex where their decline was ruthlessly sudden or painfully drawn out. In the United States we've recently seen major retail outlets go bankrupt, including the once King of retail, Sears, as well as other perineal giants -- Enron, Compaq, E.F. Hutton, and Bear Stearns. In South Korea, too, the major shipping company Hanjin went bankrupt while the auto manufacturer General Motors Korea, remains on life support. The point, of course, is not that academic disciplines are subject to the same pressures as are major industries but that the arc of success can stop, sometimes suddenly, unless problems that expose the organization to risk can be mitigated or surmounted. Progress, in other words, is not guaranteed.

In this talk I will identify two interrelated risks to our continued expansion. The first is a set of practical or procedural issues that have become institutionalized in our field and in others. Collectively, these issues are embedded in a broader system that criminologists operate in, are affected by, and respond to. This system is rooted in incentives and disincentives that, when aligned, can induce excellent science—science that is accurate, reliable, and replicable. When misaligned, however, the combination of incentives and disincentives can propel us away from rigorous and replicable science and into the land where falsehoods are embraced and touted as obviously correct. Let me suggest that we are not too far off from the latter and that an uncomfortable number of criminologists have already made that transition.

The second risk is one of intellectual culture, that if not addressed and changed will neutralize any gains made by altering the procedural issues I'll identify. Intellectual culture is a nebulous concept but what I'm referring to here is the collective willingness of our discipline to embrace the highest principles of science. Merton (1942), identified four: communism, or the sharing of ideas, information, and findings; disinterestedness (or objectivity), universalism, and organized skepticism (Macfarlane & Cheng, 2008) . Richard Feynman (1985), the famous physicist, summarized these principles as "a kind of scientific integrit y…… that corresponds to a kind of utter honesty" (p. 311). For Feynman, "utter honesty" involved the meticulous reporting of anything that could invalidate your study, as well as embracing contradictory findings that may invalidate your theory (National Academy of Sciences, 1992). Before Merton or Feynman, however, the philosopher Friedrich Nietzsche defined and discussed "intellectual honesty" and, more specifically, "intellectual conscience." For Nietzsche, the "will to knowledge" involved the scrupulous exercise of logic and judgement in the pursuit of evidence that could, or may not, lead one to a belief. Nietzsche foreshadowed much of what cognitive psychology now tells us about the formation and continuation of beliefs – namely that beliefs that bestow benefits are more likely to be formed independent of evidence, that, as Jenkins (2012, p. 268) states, "our worldview is composed of "untruths" – firmly held beliefs for which our evidence is radically inadequate." These untruths, Nietzsche argued, and science now confirms, "shape our tendency to form and evaluate new beliefs" (Jenkins, 2012, p.268). Untruths, he thus concluded, are a "condition of life." Nietzsche would love today's obsession with "fake news!"

With Nietzsche's warnings in mind, what happens when researchers embrace untruths, or when entire disciplines "sacralize," as Jonathan Haidt calls it, broad areas of study—walling them off from inquiry and attacking those who violate the sacred boundaries? What then? And what happens when scholars fail to embrace the highest principles of science, namely transparency, objectivity, and Feynman's "utter honesty?" What happens when the incentives of our scientific enterprise get misaligned and promote untruths and shoddy science? This is the question of culture I'll attempt to address as I believe it is far more pernicious than matters of methodology.

ISSUE ONE: METHODS, MAYHEM, AND REPRODUCIBILITY

All of us here owe a debt of gratitude to a psychologist named Daryl Bem. Professor Bem, from Cornell University, took eight years, nine experiments, and 1,000 subjects to show that humans were capable of precognition— yes, ESP (extra sensory perception) or knowing the future (Lowery, 2010). According to Bem, the odds that eight of his nine studies could be due to chance were 74 billion to 1. His results were published in the peer reviewed Journal of Personality and Social Psychology.

Bem's work was greeted with skepticism and a flash of dread. Nobody accused Bem of research fraud. His methods were standard experimental science in psychology and his work adhered to the basic precepts of science. Yet there it was: A study comporting to scientific standards showing something physically impossible. The implications were immediately clear: If Bem's study produce results that were not possible, how many other studies, employing equally or more rigorous designs, also produced incorrect results. LeBel and Peters (2011, p. 371) summed up the problem Bem's work posed for psychology:

> Bem (2011) deserves praise for his commitment to experimental rigor and the clarity with which he reports procedures and analyses, which generally exceed the standards of MRP (modal research practices) in empirical psychology. That being said, it is precisely because Bem's report is of objectively high quality that it is diagnostic of potential problems with MRP. By using accepted standards for experimental, analytic, and data reporting practices, yet arriving at a fantastic conclusion, Bem has put empirical psychologists in a difficult position: forced to consider either revising beliefs about the fundamental nature of time and causality or revising beliefs about the soundness of MRP.

Perhaps because psychologist make terrible theoretical physicists, most chose to revise their beliefs about the soundness of their scientific practices—practices that often include the use of experimental designs. Hence, the replication crisis was born not out of fraud or malfeasance, although psychology has suffered both, but by the faithful application of the scientific method. The story is rich in irony but there were voices prior to Bem calling for reform. One of those voices was John Ioannidis (2005) who, in a masterpiece of organized skepticism, boldly proclaimed that most published research findings were false. Ioannidis offered six corollaries to guide scholars on the likelihood findings in any one area were true. Consider his corollaries both as setting the stage for future replication efforts and for what they mean for criminology:

1: The smaller the studies conducted in a scientific field, the less likely the research findings are to be true;
2: The smaller the effect sizes in a scientific field, the less likely the research findings are to be true;
3: The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely research findings are to be true;
4: The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely research findings are to be true;
5: The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true;
6: The hotter a scientific field, the less likely the research findings are to be true.

Ioannidis went on to explain that most findings in most research areas were false positives and "may often be simply accurate measures of the prevailing bias" (p. 700). To improve research quality, he suggested larger scale studies aimed at testing major concepts where the pre-study probability was already high, moving away from null hypothesis testing, and the pre-registration of studies. In other words, address scientific processes and methods. However, he also called for a change in research culture and the "curtailing of prejudices" (p. 701). He then recommended that "…large scale studies with minimal bias should be performed on research findings that are considered relatively established, to see how often they are indeed confirmed." Perhaps validating Bem's ESP, Ioannidis presciently forecasted the results, stating unequivocally "I suspect several established "classics" will fail the test."

The period since Bem's ESP study has witnessed remarkable scholarly work in the area of replication. Research teams from around the world were mobilized and, guided by Ioannidis' insights, they keenly decided to attempt to replicate all of the major studies in psychology –studies that have been the core of teaching and research in psychology for decades. And one by one, just as Ioannidis (2005) predicted a decade earlier, they fell.

The first world-wide effort to examine replication of scientific work involved 100 studies published in three psychology journals analyzed by 270 researchers. Results were disappointing. Ninety-seven percent of the original studies reported significant results, but only 36 percent of the replication studies produced significant results; less than 50 percent of original effect sizes fell within the 95 percent replication confidence interval; 38 percent of effects were classified as having replicated, but replication effect sizes were half the magnitude of those initially reported. Studies from social psychology had a higher failure rate, 74 percent, than did studies from cognitive psychology (47 percent) (OSF, 2015). The take home message was clear: Studies that formed the backbone of psychology, many that involved experimental designs, could not be replicated, and those that could had effect sizes much lower in magnitude than originally reported. So not only did studies not replicate, even if they did many were accompanied by effect sizes that made their contribution marginal.

In short order, empirical attention turned to understanding the processes that imperiled replication efforts. Few believed, at least initially, that research fraud was sufficiently pervasive to account for the lack of replication. However, scholars for some time had been warning about the various intentional and unintentional processes researchers engage in that create unreliable findings. Charles Babbage, in 1830, for instance used the analogy of a cook "cooking" data to describe the process of selective reporting of observations. Summarizing the various degrees of freedom exercised by researchers, Simmons, Nelson, and Simonsohn (2011) discussed the "undisclosed flexibility in data collection and analysis," enjoyed by scholars. Many of the terms quickly entered researcher vernacular, including p-hacking, p-harking, asterisk hunting, and data dredging (see also, Bishop, 2019; Kerr, 1998; Obeauer & Lewandowsky, 2019). Wicherts et al., (2016) followed up and further systematized the various ways researchers can influence reported results, enumerating 34 “degrees of freedom” that can occur throughout the research process.

These degrees of freedom have become better known as Questionable Research Practices (QRP) and involve everything from fraud and fabrication to manipulating data to boost p-levels. Research into QRP’s typically involve the administration of self-report surveys that contain questions specific to individual behavior and individual reports of others behavior. Other studies, however, examine official databases. Research on QRP converge on three replicated findings: First, data fraud and data fabrication appear rare. Official estimates, which are clearly downwardly biased, suggest that fabrication, fraud, and plagiarism affect less than 1 percent of studies (George & Buyse, 2015). Self-report studies also find relatively low rates of serious data fraud, typically between 1 to 2 percent (for fraud) to 7 percent for plagiarism. That said, systematic fraud can go undetected for decades and can involve dozens of published papers. Diedrik Stapel, a Dutch social psychologist who published 130 articles and 24 book chapters, for example, was found to have falsified much of his work—work, I’ll add, that was published in the top journals in the world, such as Science. When asked how he was so successful in publishing fraudulent studies, he stated simply “I told reviewers what they wanted to hear.”

Second, the prevalence of less serious QRP, however, is substantial. Here, estimates range from 30 to almost 80 percent of researchers who admit to engaging in at least 1 QRP. John, Loewenstein, & Prelec (2012), for example, surveyed over 2,000 psychologists about their use of QRP. Their results were telling: About 10 percent of psychologists admitted to data fabrication, with large majorities admitting to other questionable practices, such as not reporting all dependent measures (78%), collecting more data after the results were known (72%), selectively reporting studies that worked (67%), and excluding data after knowing the impact of doing so (62%).

Third, when asked about the behavior of their peers, researchers report widespread use of QRP, including outright fraud. Fanelli’s (2009) meta-analysis of QRP research, for example, found that 14 percent of researchers knew of colleagues who had committed serious fraud and 72 percent who engaged in QRP. Similar patterns have been found in studies of Medicine and the health sciences (George & Buyse, 2015; Gerrits, Janse, Mulyanto, van den Berg, Klazinga, & Kringos, 2016).

Of particular concern to social scientists are the practices of p-hacking and of HARKing. P-hacking involves researchers trying various combinations of statistical models until their desired results are achieved. In a sense, the key variable reached the $p < .05$ threshold which then provides justification for attempted publication. Importantly, however, readers are never told of the efforts engaged in to obtain the published findings.

Studies show that p-hacking is widespread (Head, Holman, Lanfear, Kahn, & Jennions, 2015) and in some ways appears to be standard practice, even in our field. A lessor known, but equally problematic QRP, is that of HARKing. According to Rubin (2017, p. 2), HARKing refers to “hypothesizing after the results are known.” HARKing involves researchers combing through data conducting various statistical tests until support is found for their hypotheses. If results are contrary to initial hypotheses, however, new post hoc hypotheses are created and then passed off in the research report as original. The reader is thus lead to believe the researcher confirmed their initial hypotheses. HARKing obviously produces significant findings, which journals are more likely to publish, but it also excludes falsification since the hypotheses are always confirmed. Rubin’s summary of studies into self-reported HARKing, shown below, finds that between 27 percent and 58 percent of scholars engage in this behavior, with a mean of 43 percent.

QRP appear to be engaged in with an eye towards achieving statistical significance for parameters of interest. Examination of journal publications has decidedly shown that null effects are rarely reported, especially in the social sciences. Fanelli (2009) studied over 4,600 papers published between 1990 and 2007. In the social sciences, positive results were over twice as likely to be publish than were null results—a trend that increased over time from 1990 to 2007. By the end of the study period (2007), over 90 percent of study results found in social science journals were positive. Given standard statistical thresholds, a 90 percent confirmation rate would seem highly unlikely. Clearly, we have either achieved a level of insight into complex social behavior never before known, or our studies and the systems used to vet our studies are biased.

Researchers are not stupid people, but like anyone else they respond to incentives and disincentives that can affect their career. By any measure, publishing articles, especially in high impact journals, has become the metric by which all else is judged. Graduate students hitting the job market now often have a dozen or more publications, compared to just a few publications no more than 10 years ago. Junior scholars now go up for tenure with 30, even 60 or more publications. And senior faculty can have produced hundreds of publications over their career. Publication, for all intents and purposes, has become the currency by which status is gained, wealth is increased, and value is evaluated. What this has led to is increasing expectations for the rapid accumulation of publications and for continuity in year-to-year publication rates. As our sociological brethren have found, unreasonable standards can cause people to employ alternative strategies to achieve success. The use of QRP thus becomes a rational reaction to careerist demands and, perhaps more importantly, to the demands of publishing outlets—namely that the results reported are novel, statistically significant, and tell a good story (Bishop, 2019; Young, Ioannidis, & Al-Ubaydli, 2008).

Since positive, novel findings are more likely to get published, there are few career incentives for scholars to pursue studies that may produce insignificant results. P-hacking and other QRM may thus be born out of both ignorance of scientific formalism and an accurate assessment of the conditions necessary to achieve success in publication. That said, the expectation of journal editors and reviewers have played a critical role in incentivizing the use of QRP and the resulting lack of reliability in the criminological literature base. To be blunt, I expect most published results in criminology are the product of QRP and that few studies would replicate if such attempts were made. We are no different in this respect than are other disciplines.

The almost exclusive reliance on reaching arbitrary statistical thresholds, combined with the widespread use of QRP, is both a response to and an effect of various publication biases. I've already mention a few of these biases, such as the strong preference for significant and novel findings, but there are others. Editors often have their own views of what constitutes good science, and sometimes these views don't actually mirror good science. And as anyone who has published can tell you, editors can either kill or smooth the path for a paper to be published simply by selecting specific reviewers. Reviewers, too, sometimes have their own agendas and while I'm certain most attempt to be neutral inquisitors, it is also clear many are not. Peer review is imperfect and subject to many forms of bias. These issues were empirically examined by Gerber and Malhotra (2008), who studied 3 years of publications in the American Sociological Review, the American Journal of Sociology, and The Sociological Quarterly. Using a “caliper test,” Gerber and Malhotra found strong evidence of publication bias across all three journals. Indeed, the chance of obtaining the distribution of statistically significant results culled from these journals exceeded 1:15,000 to 1:100,000 depending on the cutoff imposed. Publication bias distorts science by providing a false or misleading picture of scientific findings. Sometimes this distortion creates an illusion of scientific consensus on an issue, while at other times the absence of null results is taken as evidence they don’t in fact exist. Either way, science becomes more illusory and misleading and scientific correction becomes less probable (Ioannidis, 2012).

Thus far I’ve imported much of my critique from research in psychology. A reasonable critic might ask whether we have a replication problem in the social science? A group of 24 scholars attempted to replicate social science experiments published in the journals Nature and Science between 2010 and 2015 (Camerer et al., 2018). Similar to the earlier OSF study on replication, this research team could only replicate 13 of the 21 original studies, with replication rates ranging from 57 to 67 percent. Effect sizes, too, were approximately ½ of those reported in initial studies. The authors argued that the presence of false positives combined with inflated effect sizes of true positives, contributed to replication failures. Combined, however, the results show that even with randomized experimental trials, from studies published in the top journals in the world, the chance for successful replication was not much better than a flip of a coin.

ISSUE TWO: DELIBERATE IGNORANCE BETRAYS SCIENTIFIC INTEGRITY

The problems I just discussed reflect deviations from the scientific process. Their fix, which I’ll propose shortly, unsurprisingly involves changing our methods and research processes to better reflect fidelity to the scientific method. What I wish to discuss now, however, has less to do with method and measurement and more to do with the embrace of scientific principles. The embrace of scientific principles seems, at least to me, to be the precondition for effective reform of our scientific processes. If we cannot embrace the most fundamental of scientific values, or if we embrace them only situationally, then changes in processes will be mute.

Now, too, seems an ideal time to discuss just how well we embrace core scientific values. Criminology, after all, is facing a crisis of legitimacy and, like many such crises, the warning signs have been visible for some time. Take, for example, the current handling of allegations of research impropriety made by a coauthor of a research team—allegations that affect a broad swath of papers published in top ranked journals and allegations that have now spilled outside of the field. Let me emphasize that I have nothing against the authors or other individuals involved in this complex drama. I do not envy any of their experiences. Nonetheless, it is fair to say that every mistake that could be made in handling this issue has been made, and that it is nothing short of astonishing how poorly these allegations have been managed. The comedy of errors has been an embarrassment to our discipline and, unfortunately, it appears as though every effort is being made to either avoid acting on the allegations or to simply sweep them away.

Accusations of research malfeasance, especially of data fabrication, are the most serious that can be leveled at a scholar. The mere accusation has the ability to forever taint one's career. However, once made two processes should kick into action—both of which are rooted in scientific values. First, in keeping with the highest principles of science, the accused should make every effort to solve the issue by providing access to the data in question. In situations where special conditions apply to the data, such as confidentiality requirements, alternative mechanisms can be arranged. Errors, if made, can be claimed and the scientific record corrected. Second, if the allegations cannot be resolved, innocence must still be presumed and all due process rights protected, but the allegations should still be adjudicated by an impartial panel of experts and the papers in question noted by the journals involved. The adjudicatory process should be guided by the principles of impartiality and objectivity.

Unfortunately, these principles gave way to collective self-interest, where each actor took steps to shield themselves or others or to adjudicate the motives of each other in public. The primary scientific questions concerning the accuracy and reliability of published research results were treated as a tertiary issue of little import. Indeed, the editor of Criminology admitted that other “gibberish” had been published in the journal and that nothing was done. Even being charitable, I find it difficult to defend the cavalier disregard for scientific accuracy and integrity. The eventual retraction of four papers, with the potential for others looming on the horizon, did not resolve these issues.

Again, my intention is not to cast aspersions at individual actors but to situate their actions in a broader context of institutional incentives and constraints—incentives and constraints that can easily become misaligned away from the values of science. If we valued transparency, for example, we would be able to examine the processes that led to so many papers being published in top journals without reviewers or editors catching some fairly obvious problems. We would know if the errors were caught and explained away, who reviewed these papers, and whether the reviews were sufficient. In short, we would know why the papers were accepted for publication by the editors and whether correctable errors were made. An emphasis on the scientific value of transparency would allow answers to these questions. After all, a good faith effort may have been made by all involved.

Transparency, objectivity, and ruthless honesty are guiding scientific values that have proven, over many generations, to lead to better science. Scientific values matter, and like Bishop (2019, p. 3), it is important that we “understand the mechanisms that maintain bad practices in individual humans. Bad science,” she astutely notes, “is usually done because somebody mistook it for good science.” In this case, many people mistook bad science for good and we might want to know why. Perhaps, however, we don’t want to know why and instead wish to remain deliberately ignorant?

Before you dismiss my comment as that of a cynic, know that deliberate ignorance is often times a rational, even desirable, state. Hertwig and Engel (2016), for example, tell us that deliberate ignorance is often preferred because it increases regret avoidance, because it can be performance enhancing, and because it can be used strategically to avoid responsibility and liability. Deliberate ignorance is also often perceived to increase impartiality and to help us maintain a range of preferred beliefs. Deliberate ignorance is, in many ways, a sensible short-term response to information that may be accompanied by psychological and emotional burdens. Not knowing, in other words, excuses our obligation to change in light of new information.

Of course, deliberate ignorance is contrary to the aims of science. Yet here too, I wish to point out that criminology has elected to remain willfully ignorant as a science. As some of you know, much of my work has been in an area called biosocial criminology. It's an area interested in how human biological variation and functioning affects human conduct. The area is more of a paradigm than a theory so many different methodological designs are employed, often from disciplines outside of criminology. One design is that of a twin study where standard quantitative genetics models are used to estimate how much variance in a trait or behavior can be attributed to unknown genetic influences, and common or unique environmental influences. Twin studies are used widely across disciplines as diverse as agriculture, to animal breeding, to brain studies, to studies of complex traits. Thousands of twin studies exist and they have yielded important insights into the origins and plasticity of human functioning and disease. Indeed, so consistently replicated are twin studies into human traits and behaviors that today it is common knowledge that all traits and behaviors are heritable, to varying degrees, and that unique environmental experiences account for more variance than do shared environments. These, by the way, are referred to as the Three Laws of Behavioral Genetics (Turkheimer, 2000).

Perhaps I exaggerated slightly when I said behavioral genetic findings were common knowledge. They are common knowledge in many sciences but not in criminology. Despite reams of replicated evidence, criminology has remained defiantly ignorant of research in this area. Let me explain: Name another area in criminology, for example, where journal editors would brag publicly about teaching their students to "hate read" specific scholarly studies, or another area where journals have banned the use of a national dataset because it was often employed by specific researchers, or where journal editors colluded to reject submissions from a specific academic area? You would be hard pressed to find such reactions. However, to better highlight the discipline's intellectual counter efforts, see if you can name any other area where critics would openly advocate banning research while simultaneously suggesting politically correct ways of discussing specific research findings. Now imagine those efforts were published in our top journal. I am, of course, referring to an exchange we had between

Professors Burt and Simmons (2014, 2015; see also, Barnes et al., 2014 and Wright et al., 2015) in the pages of Criminology. Burt and Simmons not only called for banning quantitative genetic models on grounds that they were “fatally flawed.” I will be blunt. Their piece was factually wrong in almost every way imaginable, and had their criticisms been correct, they would have upended decades of research in multiple hard sciences while simultaneously calling into question everything we know about the mathematics of evolution. Their work would have been so revolutionary, they would have earned a Nobel Prize-- had they been correct.

THE WAY FORWARD

Having exhausted my time, and I’m sure your patience, let me quickly outline a few suggestions for reform. While my talk has been critical of our field, I’m also cautiously optimistic. My optimism springs not from a naïve belief that change will be easy but from the belief that change will be hard and challenging, yet worthwhile, and I believe that most criminologists, especially younger criminologists, are interested in change. Other fields, too, have faced these daunting challenges and can they provide us with keen insights into what will most likely work for us, and what will not. Looking at these fields, many are moving to an open science framework. While details vary, the general principle is that every effort is made to make available data and statistical code so that others can easily evaluate and replicate our analytical efforts. Some journals now require data and code to be deposited prior to publication, or for authors to explain why such arrangements are not possible. Other fields have also moved to a system of preregistered studies. Preregistration is an effort to compel scholars to more clearly think about their study design, selection of variables, and planned analytical techniques prior to engaging in the study. Preregistration is designed to reduce QRP and in at least one study has been shown to dramatically reduce the number of significant associations reported in clinical trials (from 57% prior to 2000, to only 8% after year 2000) (Kaplin & Irvin, 2015).

In their “manifesto for reproducible science,” Munafo et al. (2017) recommend 10 proposals to systematize the collection and reporting of social scientific data. Their recommendations run the gambit from rewarding scholars who participate in open science efforts, to creating and using protocol checklists for data reporting, to engaging in collaborative and team research efforts. If taken seriously, criminology could be improved by embracing these, and similar, practices to make our science more transparent and hopefully more reliable. Yes, retractions may increase but as others have noted, retractions are a sign of a healthy science (Fanelli, 2013). There is little reason why criminology should avoid moving in a similar direction.

Open science, however, is not a cure all for what ails our discipline. In recent years our major organizations and organization presidents have encouraged scholars to engage in political activism. The scholar-activist model they propose couples the passions for social and economic justice to scholarly research efforts. Such language has now been codified in the American Society of Criminology's Code of Ethics. This is a terrible mistake because it frames the scientific process in terms of providing evidence about favored narratives so as to justify specific policies. Under this scheme, science is highjacked and made slave to the political whims of its masters. As a host of studies show, ideological reasoning impairs logical judgement and reduces the safeguards science offers. In the end, such an approach is guaranteed to delegitimate our science and to divorce our work from reality (Martin, 2015).

Criminology stands at a fork in the road. May I suggest we take the path less traveled, that we embrace Feynman's "utter honesty" and Nietzsche's "intellectual conscience," and that we open our science and confront directly the challenges that will emerge. Progress, after all, is never guaranteed.

References

Barnes, J. C., Wright, J. P., Boutwell, B. B., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., & Beaver, K. M. (2014). Demonstrating the validity of twin research in criminology. *Criminology*, *52*(4), 588–626.

Bishop, D. V. (2019). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, 174702181988651. https://doi.org/10.1177/1747021819886519

Burt, C. H., & Simons, R. L. (2014). Pulling back the curtain on heritability studies: Biosocial criminology in the postgenomic era. *Criminology*, *52*(2), 223–262.

Burt, C. H., & Simons, R. L. (2015). Heritability studies in the postgenomic era: The fatal flaw is conceptual. *Criminology*, *53*(1), 103–112.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Eric Turkheimer. (2000). Three Laws of Behavioral Genetics and What They Mean. *Current Directions in Psychological Science*, *9*(5), 160–164.

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *PLoS ONE*, *4*(5), e5738. https://doi.org/10.1371/journal.pone.0005738

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Fanelli, D. (2013). Why Growing Retractions Are (Mostly) a Good Sign. *PLoS Medicine*, *10*(12), e1001563. https://doi.org/10.1371/journal.pmed.1001563

George, S. L., & Buyse, M. (2015). Data fraud in clinical trials. *Clinical Investigation*, *5*(2), 161–173. https://doi.org/10.4155/cli.14.116

Gerber, A. S., & Malhotra, N. (2008). Publication Bias in Empirical Sociological Research: Do Arbitrary Significance Levels Distort Published Results? *Sociological Methods & Research*, *37*(1), 3–30. https://doi.org/10.1177/0049124108318973

Gerrits, R. G., Jansen, T., Mulyanto, J., van den Berg, M. J., Klazinga, N. S., & Kringos, D. S. (2019). Occurrence and nature of questionable research practices in the reporting of messages and conclusions in international scientific Health Services Research publications: A structured assessment of publications authored by researchers in the Netherlands. *BMJ Open*, *9*(5), e027903. https://doi.org/10.1136/bmjopen-2018-027903

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The Extent and Consequences of P-Hacking in Science. *PLOS Biology*, *13*(3), e1002106. https://doi.org/10.1371/journal.pbio.1002106

Hertwig, R., & Engel, C. (2016). Homo Ignorans: Deliberately Choosing Not to Know. *Perspectives on Psychological Science*, *11*(3), 359–372. https://doi.org/10.1177/1745691616635594

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), 6.

Ioannidis, J. P. A. (2012). Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, *7*(6), 645–654. https://doi.org/10.1177/1745691612464056

Jenkins, S. (2012). Nietzsche's Questions Concerning the Will to Truth. *Journal of the History of Philosophy*, *50*(2), 265–289. https://doi.org/10.1353/hph.2012.0030

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. *PLOS ONE*, *10*(8), e0132382. https://doi.org/10.1371/journal.pone.0132382

Klaus Oberauer, & Stephan Lewandowsky. (2019). Addressing the theory crisis in psychology. *Psychometric Bulletin & Review*, *25*(5), 1596–1618.

LeBel, E. P., & Peters, K. R. (2011). Fearing the Future of Empirical Psychology: Bem's (2011) Evidence of Psi as a Case Study of Deficiencies in Modal Research Practice. *Review of General Psychology*, *15*(4), 371–379. https://doi.org/10.1037/a0025172

Macfarlane, B., & Cheng, M. (2008). Communism, Universalism and Disinterestedness: Re-examining Contemporary Support among Academics for Merton's Scientific Norms. *Journal of Academic Ethics*, 1–12. https://doi.org/10.1007/s10805-008-9055-y

Martin, C. C. (2015). How Ideology Has Hindered Sociological Insight.

American Sociologist. https://doi.org/10.1007/s12108-015-9263-z

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1). https://doi.org/10.1038/s41562-016-0021

Norbert L. Kerr. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Rubin, M. (2017). When Does HARKing Hurt? Identifying When Different Types of Undisclosed Post Hoc Hypothesizing Harm Scientific Progress. *Review of General Psychology*, *21*(4), 308–320. https://doi.org/10.1037/gpr0000128

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01832

Wright, J. P., Barnes, J. C., Boutwell, B. B., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., & Beaver, K. M. (2015). Mathematical proof is not minutiae and irreducible complexity is not a theory: A final response to burt and simons and a call to criminologists. *Criminology*, *53*(1). https://doi.org/10.1111/1745-9125.12059

Young, N. S., Ioannidis, J. P. A., & Al-Ubaydli, O. (2008). Why Current Publication Practices May Distort Science. *PLoS Medicine*, *5*(10), e201. https://doi.org/10.1371/journal.pmed.0050201