The Political Methodologist

Newsletter of the Political Methodology Section
American Political Science Association
Volume 23, Number 1, Fall 2015

Editor:
Justin Esarey, Rice University
justin@justinesarey.com

Editorial Assistant:
Ahra Wu, Rice University
ahra.wu@rice.edu

Associate Editors:
Randolph T. Stevenson, Rice University
stevenso@rice.edu

Rick K. Wilson, Rice University
rkw@rice.edu

Contents

Notes From the Editors

The editorial team at *The Political Methodologist* is proud to present this special issue on peer review! In this issue, eight political scientists comment on their experiences as authors, reviewers, and/or editors dealing with the scientific peer review process and (in some cases) offer suggestions to improve that process. Our contributions come from authors at very different levels of rank and subfield specialization in the discipline and consequently represent a diversity of viewpoints within political methodology.

Additionally, some of the contributors to this issue will participate in an online roundtable discussion on March 18th at 12:00 noon (Eastern time) as a part of the International Methods Colloquium. If you want to add your voice to this discussion, we encourage you to join the roundtable audience! Participation in the roundtable discussion is free and open to anyone around the world (with internet access and a PC or Macintosh). Visit the IMC registration page linked here to register to participate or visit www.methods-colloquium.com for more details.

The initial stimulus for this special issue came from a discussion of the peer review process among political scientists on Twitter, and the editorial team is always open to ideas for new special issues on topics of importance to the methods community. If you have an idea for a special issue (from social media or elsewhere), feel free to let us know! We believe that special issues are a great way to encourage discussion among political scientists and garner attention for new ideas in the discipline.

The Editors

Introduction to the Special Issue: Acceptance Rates and the Aesthetics of Peer Review

Justin Esarey
Rice University
justin@justinesarey.com

Based on the contributions to *The Political Methodologist*'s special issue on peer review, it seems that many political scientists are not happy with the kind of feedback they receive from the peer review process. A theme seems to be that reviewers focus less on the scientific merits of a piece – viz., what can be learned from the evidence offered – and more on whether the piece is to the reviewer's taste in terms of questions asked and methodologies employed. While I agree that this feedback is unhelpful and undesirable, I am also concerned that it is a fundamental feature of the way our peer review system works. More specifically, I believe that a system of journals with prestige tiers enforced by extreme selectivity creates a review system where scientific soundness is a necessary but far from sufficient criteria for publication, meaning that fundamentally aesthetic and sociological factors ultimately determine what gets published and inform the content of our reviews.

As Brendan Nyhan says, "authors frequently despair not just about timeliness of the reviews they receive but their focus." Nyhan seeks to improve the focus of reviews by offering a checklist of questions that reviewers should answer as a part of their reviews (omitting those questions that, presumably, they should *not* seek to answer). These questions revolve around ensuring that evidence offered is consistent with conclusions ("Does the author control for or condition on variables that could be affected by the treatment of interest?") and that statistical inferences are unlikely to be spurious ("Are any subgroup analyses adequately powered and clearly motivated by theory rather than data mining?").

The other contributors express opinions in sync with Nyhan's point of view. For example, Tom Pepinsky says:

> "*I strive to be indifferent to concerns of the type 'if this manuscript is published, then people will work on this topic or adopt this methodology, even if I think it is boring or misleading?' Instead, I try to focus on questions like 'is this manuscript accomplishing what it sets out to accomplish?' and 'are there ways to my comments can make it better?' My goal is to judge the manuscript on its own terms.*"

Relatedly, Sara Mitchell argues that reviewers should focus on "criticisms internal to the project rather than moving to a purely external critique." This is explored more fully in the piece by Krupnikov and Levine, where they argue that simply writing "external validity concern!" next to any laboratory experiment hardly addresses whether the article's evidence actually answers the questions offered; in a way, the attitude they criticize comes uncomfortably close to arguing that any question that can be answered using laboratory experiments doesn't deserve to be asked, *ipso facto*.

My own perspective on what a peer review ought to be has changed during my career. Like Tom Pepinsky, I once thought my job was to "protect" the discipline from "bad research" (whatever that means). Now, I believe that a peer review ought to answer just one question: **What can we learn from this article?**[1]

Specifically, I think that every sentence in a review ought to be:

1. a factual statement about what the author believes can be learned from his/her research, or
2. a factual statement of what the reviewer thinks actually can be learned from the author's research, or
3. an argument about why something in particular can (or cannot) be learned from the author's research, supported by evidence.

This feedback helps an editor learn what marginal contribution that the submitted paper makes to our understanding, informing his/her judgment for publication. It also helps the author understand what s/he is communicating in his/her piece and whether claims must be trimmed or hedged to ensure congruence with the offered evidence (or more evidence must be offered to support claims that are central to the article).

Things that I think shouldn't be addressed in a review include:

1. whether the reviewer thinks the contribution is sufficiently important to be published in the journal
2. whether the reviewer thinks other questions ought to have been asked and answered
3. whether the reviewer believes that an alternative methodology would have been able to answer different or better questions
4. whether the paper comprehensively reviews extant literature on the subject (unless the paper defines itself as a literature review)

In particular, I think that the editor is the person in the most appropriate position to decide whether the contribution is sufficiently important for publication, as that is a part of his/her job; I also think that such a decision should

[1] Our snarkier readers may be thinking that this question can be answered in just one word for many papers they review: "nothing." I cannot exclude that possibility, though it is inconsistent with my own experience as a reviewer. I would say that, if a reviewer believed nothing can be learned from a paper, I would hope that the reviewer would provide feedback that is lengthy and detailed enough to justify that conclusion.

be made (whenever possible) by the editorial staff *before* reviews are solicited. (Indeed, in another article (Esarey N.d.) I offer simulation evidence that this system actually‘ produces better journal content, as evaluated by the overall population of political scientists, compared to a more reviewer-focused decision procedure.) Alternatively, the importance of a publication could be decided (as Danilo Freire alludes) by the discipline at large, as expressed in readership and citation rates, and not by one editor (or a small number of anonymous reviewers); such a system is certainly conceivable in the post-scarcity publication environment created by online publishing.

Of course, as our suite of contributions to *The Political Methodologist* makes clear, most of us do *not* receive reviews that are focused narrowly on the issues that I have outlined. Naturally, this is a frustrating experience. I think it is particularly trying to read a review that says something like, “this paper makes a sound scientific contribution to knowledge, but that contribution is not important enough to be published in journal X.” It is annoying precisely because the review acknowledges that the paper isn’t flawed, but simply isn’t to the reviewer’s taste. It is the academic equivalent of being told that the reviewer is “just not that into you.” It is a fundamentally unactionable criticism.

Unfortunately, I believe that authors are likely to receive *more*, not *less*, of such feedback in the future regardless of what I or anyone else may think. The reason is that journal acceptance rates are so low, and the proportion of manuscripts that make sound contributions to knowledge is so high, that other criteria must necessarily be used to select from those papers which will be published from the set of those that could credibly be published.

Consider that in 2014, the *American Journal of Political Science* accepted only 9.6% of submitted manuscripts (Jacoby et al. 2015) and International Studies Quarterly accepted about 14% (Nexon 2016). The trend is typically downward: at *Political Research Quarterly* (Clayton et al. 2011), acceptance rates fell by 1/3 between 2006 and 2011 (to just under 12 percent acceptance in 2011). I speculate that far more than 10-14% of the manuscripts received by *AJPS*, *ISQ*, and *PRQ* were scientifically sound contributions to political science that could have easily been published in those journals – at least, this is what editors tend to write in their rejection letters!

When (let us say, for argument’s sake) 25% of submitted articles are scientifically sound but journal acceptance rates are less than half that value, it is essentially required that editors (and, by extension, reviewers) must choose on criteria other than soundness when selecting articles for publication. It is natural that the slippery and socially-constructed criterion of “importance” in its many guises would come to the fore in such an environment. Did the paper address questions you think are the most “interesting?” Did the paper use what you believe are the most “cutting edge” methodologies? “Interesting” questions and “cutting edge” methodologies are aesthetic judgments, at least in part, and defined relative to a group of people making these aesthetic judgments. Consequently, I fear that the peer review process *must* become as much a function of sociology as of science because of the increasingly competitive nature of journal publication. Insomuch that I am correct, I think would prefer that these aesthetic judgments come from the discipline at large (as embodied in readership rates and citations) and not from two or three anonymous colleagues.

Still, as long as there are tiers of journal prestige and these tiers are a function of selectivity, I would guess that the power of aesthetic criteria to influence the peer review process has to persist. Indeed, I speculate that the proportion of sound contributions in the submission pool is trending upward because of the intensive focus of many PhD programs on rigorous research design training and the ever-increasing requirements of tenure and promotion committees. At the very least, the number of submissions is going up (from 134 in 2001 to 478 in 2014 at *ISQ*), so even if quality is stable selectivity must rise if the number of journal pages stays constant. Consequently, I fear that a currently frustrating situation is likely to get worse over time, with articles being selected for publication in the most prominent journals of our discipline on progressively more whimsical criteria.

What can be done? At the least, I think we can recognize that the “tiers” of journal prestige do not necessarily mean what they might have used to in terms of scientific quality or even interest to a broad community of political scientists and policy makers. Beyond this, I am not sure. Perhaps a system that rewards authors more for citation rates and less for the “prestige” of the publication outlet might help. But undoubtedly these systems would *also* have unanticipated and undesirable properties, and it remains to be seen whether they would truly improve scholarly satisfaction with the peer review system.

References

Clayton, Cornell W., Amy Mazur, Season Hoard, and Lucas McMillan. 2011. “Political Research Quarterly Activity Report 2011.” Presented to the Editorial Advisory Board, WPSA Annual Meeting, Portland, OR, March 22-24, 2012. `http://wpsa.research.pdx.edu/pub/PRQ/prq2011-2012report.pdf` (February 1, 2016).

Esarey, Justin. N.d. “How Does Peer Review Shape Science? A Simulation Study of Editors, Reviewers, and the Scientific Publication Process.” Unpublished Manuscript. `http://jee3.web.rice.edu/peer-review.pdf` (February 1, 2016).

Jacoby, William G., Robert N. Lupton, Miles T. Armaly, Adam Enders, Marina Carabellese, and Michael Sayre. 2015. “American Journal of Political Science: Report to the Editorial Board and the Midwest Political Science Association Executive Council.” April.

https://ajpsblogging.files.wordpress.com/2015/04/ajps-editors-report-on-2014.pdf (February 1, 2016).

Nexon, Daniel H. 2016. "ISQ Annual Report, 2015." January 5. https://dl.dropboxusercontent.com/u/22744489/ISQ%202015%20report.pdf (February 1, 2016).

A Checklist Manifesto for Peer Review

Brendan Nyhan
Dartmouth College
nyhan@dartmouth.edu

The problems with peer review are increasingly recognized across the scientific community. Failures to provide timely reviews often lead to interminable delays for authors, especially when editors force authors to endure multiple rounds of review (e.g., Smith 2014). Other scholars simply refuse to contribute reviews of their own, which recently prompted the *AJPS* editor to propose a rule stating that he "reserves the right to refuse submissions from authors who repeatedly fail to provide reviews for the Journal when invited to do so" (Jacoby 2015).

Concerns over delays in the publication process have prompted a series of proposals intended to improve the peer review system. Diana Mutz and I recently outlined how frequent flier-type systems might improve on the status quo by rewarding scholars who provide rapid, high-quality reviews (Mutz 2015; Nyhan 2015). Similarly, Chetty, Saez, and Sándor (2014) report favorable results from an experiment testing the effects of requesting reviews on shorter timelines, promising to publish reviewer turnaround times, and offering financial incentives.

While these efforts are worthy, their primary goal is to speed the review process and reduce the frequency that scholars decline to participate rather than to improve the reviews that journals receive. However, there are also reasons for concern about the value of the content of reviews under the status quo, especially given heterogeneous definitions of quality among reviewers. Esarey (N.d.), for instance, uses simulations to show that it is unclear to what extent (if at all) reviews help select the best articles. Similarly, Price (2014) found only modest overlap in the papers selected for a computer science conference when they were assigned to two sets of reviewers (see also Mahoney 1977).

More generally, authors frequently despair not just about timeliness of the reviews they receive but their focus. Researchers often report that reviewers tend to focus on the way that articles are framed (e.g., Cohen 2015; Lindner 2015). These anecdotes are consistent with the findings of Goodman et al. (1994), who estimate that three of the five areas where medical manuscripts showed statistically significant improvements in quality after peer review were related to framing ("discussion of limitations," "acknowledgement and justification of generalizations," and "appropriateness of the strength or tone of the conclusions"). While sometimes valuable, these suggestions are largely aesthetic and can often bloat published articles, especially in social science journals with higher word limits. Useful suggestions for improving measurement or statistical analyses are seemingly more rare. One study found, for instance, that authors were most likely to cite their discussion sections as having been improved by peer review; methodology and statistics were much less likely to be cited (Mulligan, Hall, and Raphael 2013).

Why not try to shift the focus of reviews in a more valuable direction? I propose that journals try to nudge reviewers to focus on areas where they can most effectively improve the scientific quality of the manuscript under consideration using checklists, which are being adopted in medicine after widespread use in aviation and other fields (e.g., Haynes et al. 2009; Gawande 2009). In this case, reviewers would be asked to check off a set of yes or no items indicating that they had assessed whether both the manuscript and their review meet a set of specific standards before their review would be complete. (Any yes answers would prompt the journal website to ask the reviewer to elaborate further.) This process could help bring the quality standards of reviews into closer alignment.

The checklist items listed below have two main goals. First, they seek to reduce the disproportionate focus on framing, minimize demands on authors to include reviewer-appeasing citations, and deter unreasonable reviewer requests. Second, the content of the checklists seeks to cue reviewers to identify and correct recurring statistical errors and to also remind them to avoid introducing such mistakes in their own feedback. Though this process might add a bit more time to the review process, the resulting improvement in review quality could be significant.

Manuscript checklist:

- Does the author properly interpret any interaction terms and include the necessary interactions to test differences in relationships between subsamples? (Brambor, Clark, and Golder 2006)

- Does the author interpret a null finding as evidence that the true effect equals 0 or otherwise misinterpret p-values and/or confidence intervals? (Gill 1999)

- Does the author provide their questionnaire and any other materials necessary to replicate the study in an appendix?
- Does the author use causal language to describe a correlational finding?
- Does the author specify the assumptions necessary to interpret their findings as causal?
- Does the author properly specify and test a mediation model if one is proposed using current best practices? (Imai, Keele, and Tingley 2010)
- Does the author control for or condition on variables that could be affected by the treatment of interest? (Rosenbaum 1984; Elwert and Winship 2014)
- Does the author have sufficient statistical power to test the hypothesis of interest reliably? (Simmons 2014)
- Are any subgroup analyses adequately powered and clearly motivated by theory rather than data mining?

Review checklist:

- Did you request that a control variable be included in a statistical model without specifying how it would confound the author's proposed causal inference?
- Did you request any sample restrictions or control variables that would induce post-treatment bias? (Rosenbaum 2014; Elwert and Winship 2014)
- Did you request a citation to or discussion of an article without explaining why it is essential to the author's argument?
- Could the author's literature review be shortened? Please justify specifically why any additions are required and note areas where corresponding reductions could be made elsewhere.
- Are your comments about the article's framing relevant to the scientific validity of the paper? How specifically?
- Did you request a replication study? Why is one necessary given the costs to the author?
- Does your review include any unnecessary or ad hominem criticism of the author?
- If you are concerned about whether the sample is sufficiently general or representative, did you provide specific reasons why the author's results would not generalize and/or propose a feasible design that would overcome these limitations? (Druckman and Kam 2011; Aronow and Samii 2015)
- Do your comments penalize the authors in any way for null findings or suggest ways they can find a significant relationship?
- If your review is positive, did you explain the contributions of the article and the reasons you think it is worthy of publication in sufficient detail? (Nexon 2015)

Acknowledgments

I thank John Carey, Justin Esarey, Jacob Montgomery, and Thomas Zeitzoff for helpful comments.

References

Aronow, Peter M. and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250-67.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63-82.

Chetty, Raj, Emmanuel Saez, and László Sándor. 2014. "What Policies Increase Prosocial Behavior? An Experiment with Referees at the Journal of Public Economics." *Journal of Economic Perspectives* 28 (3): 169-88.

Cohen, Philip N. 2015. "Our Broken Peer Review System, in One Saga." October 5. `https://familyinequality.wordpress.com/2015/10/05/our-broken-peer-review-system-in-one-saga/` (February 1, 2016).

Druckman, James N., and Cindy D. Kam. 2011. "Students as Experimental Participants: A Defense of the "Narrow Data Base."" In *Cambridge Handbook of Experimental Political Science*, eds. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. Cambridge: Cambridge University Press, 41-57.

Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40: 31-53.

Esarey, Justin. N.d. "How Does Peer Review Shape Science? A Simulation Study of Editors, Reviewers, and the Scientific Publication Process." Unpublished Manuscript. `http://jee3.web.rice.edu/peer-review.pdf` (February 1, 2016).

Gawande, Atul. 2009. *The Checklist Manifesto: How to Get Things Right.* New York: Metropolitan Books.

Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52 (3): 647-74.

Goodman, Steven N., Jesse Berlin, Suzanne W. Fletcher, and Robert H. Fletcher. 1994. "Manuscript Quality before and after Peer Review and Editing at *Annals of Internal Medicine*." *Annals of Internal Medicine* 121 (1): 11-21.

Haynes, Alex B., Thomas G. Weiser, William R. Berry, Stuart R. Lipsitz, Abdel-Hadi S. Breizat, E. Patchen Dellinger, Teodoro Herbosa, Sudhir Joseph, Pascience L. Kibatala, Marie Carmela M. Lapitan, Alan F. Merry, Krishna Moorthy, Richard K. Reznick, Bryce Taylor, and Atul A. Gawande, for the Safe Surgery Saves Lives Study Group. 2009. "A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population." *New England Journal of Medicine* 360 (5): 491-9.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15 (4): 309-34.

Jacoby, William G. 2015. "Editor's Midterm Report to the Executive Council of the Midwest Political Science Association." August 27. `https://ajpsblogging.files.wordpress.com/2015/09/ajps-2015-midyear-report-8-27-15.pdf` (February 1, 2016).

Lindner, Andrew M. 2015. "How the Internet Scooped Science (and What Science Still Has to Offer)." March 17. `https://academics.skidmore.edu/blogs/alindner/2015/03/17/how-the-internet-scooped-science-and-what-science-still-has-to-offer/` (February 1, 2016).

Mahoney, Michael J. 1977. "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System." *Cognitive Therapy and Research* 1 (2): 161-75.

Mulligan, Adrian, Louise Hall, and Ellen Raphael. 2013. "Peer Review in a Changing World: An International Study Measuring the Attitudes of Researchers." *Journal of the American Society for Information Science and Technology* 64 (1): 132-61.

Mutz, Diana C. 2015. "Incentivizing the Manuscript-review System Using REX." *PS: Political Science & Politics* 48 (S1): 73-7.

Nexon, Daniel. 2015. "Notes for Reviewers: The Pitfalls of a First Round Rave Review." November 4. `http://www.isanet.org/Publications/ISQ/Posts/ID/4918/Notes-for-Reviewers-The-Pitfalls-of-a-First-Round-Rave-Review` (February 1, 2016).

Nyhan, Brendan. 2015. "Increasing the Credibility of Political Science Research: A Proposal for Journal Reforms." *PS: Political Science & Politics* 48 (S1): 78-83.

Price, Eric. 2014. "The NIPS Experiment." December 15. `http://blog.mrtz.org/2014/12/15/the-nips-experiment.html` (February 1, 2016).

Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment." *Journal of the Royal Statistical Society: Series A (General)* 147 (5): 656-66.

Simmons, Joe. 2014. "MTurk vs. the Lab: Either Way We Need Big Samples." April 4, 2014. `http://datacolada.org/2014/04/04/18-mturk-vs-the-lab-either-way-we-need-big-samples/` (February 1, 2016).

Smith, Robert Courtney. 2014. "Black Mexicans, Conjunctural Ethnicity, and Operating Identities Long-Term Ethnographic Analysis." *American Sociological Review* 79 (3): 517-48.

Peering at Open Peer Review

Danilo Freire
King's College London
danilofreire@gmail.com

Introduction

Peer review is an essential part of the modern scientific process. Sending manuscripts for others to scrutinize is such a widespread practice in academia that its importance cannot be overstated. Since the late eighteenth century, when the *Philosophical Transactions of the Royal Society* pioneered editorial review,[1] virtually every scholarly outlet has adopted some sort of pre-publication assessment of received works. Although the specifics may vary, the procedure has remained largely the same since its inception: submit, receive anonymous criticism, revise, restart the process if required. A recent survey of APSA members indicates that political scientists overwhelmingly believe in the value of peer review (95%) and the vast majority of them (80%) think peer review is a useful tool to keep themselves up to date with cutting-edge research (Djupe 2015, 349). But do these figures suggest that journal editors can rest upon their laurels and leave the system as it is?

[1]Several authors affirm that the first publication to implement a peer review system similar to what we have today was the *Medical Essays and Observations*, edited by the Royal Society of Edinburgh 1731 (Fitzpatrick 2011; Kronick 1990; Lee et al. 2013). However, the current format of "one editor and two referees" is surprisingly recent and was adopted only after the Second World War (Rowland 2002; Weller 2001, 3-8). An early predecessor to peer review was the Arab physician Ishaq ibn Ali Al-Ruhawi (CE 854 – 931), who argued that physicians should have their notes evaluated by their peers and, eventually, be sued if the reviews were unfavorable (Spier 2002, 357). Fortunately, his last recommendation has not been strictly enforced in our times.

Not quite. A number of studies have been written about the shortcomings of peer review. The system has been criticized for being too slow (Ware 2008), conservative (Eisenhart 2002), inconsistent (Smith 2006; Hojat, Gonnella, and Caelleigh 2003), nepotist (Sandström and Hällsten 2008), biased against women (Wennerås and Wold 1997), affiliation (Peters and Ceci 1982), nationality (Ernst and Kienbacher 1991) and language (Ross et al. 2006). These complains have fostered interesting academic debates (e.g. Meadows 1998; Weller 2001), but thus far the literature offers little practical advice on how to tackle peer review problems. One often overlooked aspect in these discussions is how to provide incentives for reviewers to write well-balanced reports. On the one hand, it is not uncommon for reviewers to feel that their work is burdensome and not properly acknowledged. Further, due to the anonymous nature of the reviewing process itself, it is impossible to give the referee proper credit for a constructive report. On the other hand, the reviewers' right to full anonymity may lead to sub-optimal outcomes as referees can rarely be held accountable for being judgemental (Fabiato 1994).

In this short article, I argue that open peer review can address these issues in a variety of ways. Open peer review consists in requiring referees to sign their reports and requesting editors to publish the reviews alongside the final manuscripts (DeCoursey 2006). Additionally, I suggest that political scientists would benefit from using an existing, or creating a dedicated, online repository to store referee reports relevant to the discipline. Although these ideas have long been implemented in the natural sciences (DeCoursey 2006; Ford 2013; Ford 2015; Greaves et al. 2006; Pöschl 2012), the pros and cons of open peer review have rarely been discussed in our field. As I point out below, open peer reviews should be the norm in the social sciences for numerous reasons. Open peer review provides context to published manuscripts, encourages post-publication discussion and, most importantly, makes the whole editorial process more transparent. Public reports have important pedagogical benefits, offering students a first-hand experience of what are the best reviewing practices in their field. Finally, open peer review not only allows referees to showcase their expert knowledge, but also creates an additional incentive for scholars to write timely and thoughtful critiques. Since online storage costs are currently low and Digital Object Identifiers[2] (DOI) are easy to obtain, the ideas proposed here are feasible and can be promptly implemented.

In the next section, I present the argument for an open peer review system in further detail. I comment on some of the questions referees may have and why they should not be reticent to share their work. Then I show how scholars can use Publons,[3] a startup company created for this purpose, to publicize their reviews. The last section offers some concluding remarks.

Opening Yet Another Black Box

Over the last few decades, political scientists have pushed for higher standards of reproducibility in the discipline. Proposals to increase openness in the field have often sparked controversy,[4] but they have achieved considerable success in their task. In comparison to just a few years ago, data sets are widely available online,[5] open source software such as R and Python are increasingly popular in classrooms, and even version control has been making inroads into scholarly work (Gandrud 2013a; Gandrud 2013b; Jones 2013).

Open peer review (henceforth OPR) is largely in line with this trend towards a more transparent political science. Several definitions of OPR have been suggested, including more radical ones such as allowing anyone to write pre-publication reviews (crowdsourcing) or by fully replacing peer review with post-publication comments (Ford 2013). However, I believe that by adopting a narrow definition of OPR – only asking referees to sign their reports – we can better accommodate positive aspects of traditional peer review, such as author blinding, into an open framework. Hence, in this text OPR is understood as a reviewing method where both referee information and their reports are disclosed to the public, while the authors' identities are not known to the reviewers before manuscript publication.

How exactly would OPR increase transparency in political science? As noted by a number of articles on the topic, OPR creates incentives for referees to write insightful reports, or at least it has no adverse impact over the quality of reviews (DeCoursey 2006; Godlee 2002; Groves 2010; Pöschl 2012; Shanahan and Olsen 2014). In a study that used randomized trials to assess the effect of OPR in the *British Journal of Psychiatry*, Walsh et al. (2000) show that "signed reviews were of higher quality, were more courteous and took longer to complete than unsigned reviews." Similar results were reported by McNutt et al. (1990, 1374), who affirm that "editors graded signers as more constructive and courteous [...], [and] authors graded signers as fairer." In the same vein, Kowalczuk et al. (2013) measured the difference in review quality in *BMC Microbiology* and *BMC Infectious Diseases* and stated that signers received higher ratings for their feedback on methods and for the amount of evidence they mobilized to substantiate their decisions. Van Rooyen and her colleagues (1999; 2010) also ran two randomized studies on the subject, and although

[2]https://www.doi.org/ (February 1, 2016)

[3]http://publons.com/ (February 1, 2016)

[4]See, for instance, the debate over replication that followed King (1995) and the ongoing discussion on data access and research transparency (DA-RT) guidelines (Lupia and Elman 2014).

[5]As of November 16, 2015, there were about 60,000 data sets hosted on the Dataverse Network. See: http://dataverse.org/ (February 1, 2016)

they did not find a major difference in perceived quality of both types of review, they reported that reviewers in the treatment group also took significantly more time to evaluate the manuscripts in comparison with the control group. They also note authors broadly favored the open system against closed peer review.

Another advantage of OPR is that it offers a clear way for referees to highlight their specialized knowledge. When reviews are signed, referees are able to receive credit for their important, yet virtually unsung, academic contributions. Instead of just having a rather vague "service to profession" section in their CVs, referees can precise information about the topics they are knowledgeable about and which sort of advice they are giving to prospective authors. Moreover, reports assigned a DOI number can be shared as any other piece of scholarly work, which leads to an increase in the body of knowledge of our discipline and a higher number of citations to referees. In this sense, signed reviews can also be useful for universities and funding bodies. It is an additional method to assess the expert knowledge of a prospective candidate. As supervising skills are somewhat difficult to measure, signed reviews are a good proxy for an applicant's teaching abilities.

OPR provides background to manuscripts at the time of publication (Ford 2015; Lipworth et al. 2011). It is not uncommon for a manuscript to take months, or even years, to be published in a peer-reviewed journal. In the meantime, the text usually undergoes several major revisions, but readers rarely, if ever, see this trial-and-error approach in action. With public reviews, everyone would be able to track the changes made in the original manuscript and understand how the referees improved the text before its final version. Hence, OPR makes the scientific exchange clear, provides useful background information to manuscripts and fosters post-publication discussions by the readership at large.

Signed and public reviews are also important pedagogical tools. OPR gives a rare glimpse of how academic research is actually conducted, making explicit the usual need for multiple iterations between the authors and the editors before an article appears in print. Furthermore, OPR can fill some of the gap in peer-review training for graduate students. OPR allows junior scholars to compare different review styles, understand what the current empirical or theoretical puzzles of their discipline are, and engage in post-publication discussions about topics in which they are interested (Ford 2015; Lipworth et al. 2011).

One may question the importance of OPR by affirming, as does Khan (2010), that "open review can cause reviewers to blunt their opinions for fear of causing offence and so produce poorer reviews." The problem would be particularly acute for junior scholars, who would refrain from offering honest criticism to senior faculty members due to fear of reprisals (Wendler and Miller 2014, 698). This argument, however, seems misguided. First, as noted above, thus far there is no empirical evidence that OPR is detrimental to the quality of reviews. Therefore, we can easily turn this criticism upside down and suggest it is not OPR that needs to justify itself, but rather the traditional system (Godlee 2002; Rennie 1998). Since closed peer reviews do not lead to higher-quality critiques, one may reasonably ask why should OPR *not* be widely implemented on ethical grounds alone. In addition, recent replication papers by young political scientists have been successful at pointing out mistakes in scholarly work of seasoned researchers (e.g. Bell and Miller 2015; Broockman, Kalla, and Aronow 2015; McNutt 2015). These replications have been specially helpful for graduate students building their careers (King 2006), and *a priori* there is no reason why insightful peer reviews could not have the same positive effect on a young scholar's academic reputation.

A second criticism of OPR, closely related to the first, says it causes higher acceptance rates because reviewers become more lenient in their comments. While there is indeed some evidence that reviewers who opt for signed reports are slightly more likely to recommend publication (Walsh et al. 2000), the concern over excessive acceptance rates seems unfounded. First, higher acceptance can be a *positive* outcome if it reflects the difference between truly constructive reviews and overly zealous criticisms from the closed system (Fabiato 1994, 1136). Moreover, if a manuscript is deemed relevant by editors and peers, there is no reason why it should not eventually be accepted. Further, if the reviews are made available, the whole publication process can be verified and contested if necessary. There is no need to be reticent about low-quality texts making the pages of flagship journals.

Implementation

Open peer reviews are intended to be public by design, thus it is important to devote some thought to the best ways to make report information available. Since a few science journals have already implemented variants of OPR, they are a natural starting point for our discussion. The *BMJ*[6] follows a simple yet effective method to publicize its reports: all accompanying texts are available alongside the published article in a tab named "peer review." The reader can access related reviews and additional materials without leaving the main text page, which is indeed convenient. Similar methods have also been adopted by open access journals such as

[6] http://www.bmj.com/ (February 1, 2016)
[7] http://f1000research.com/ (February 1, 2016)
[8] http://peerj.com/ (February 1, 2016)
[9] http://rsos.royalsocietypublishing.org/ (February 1, 2016)

F1000Research,[7] *PeerJ*[8] and *Royal Society Open Science*.[9] Most publications let authors decide whether reviewing information should be available to the public. If a researcher does not feel comfortable sharing the comments they receive, he or she can opt out of OPR simply by notifying the editors. In the case of *BMJ* and *F1000Research*, however, OPR is mandatory.

In this regard, editors must weigh the pros and cons of making OPR compulsory. While OPR is to be encouraged for the reasons stated above, it is also crucial that authors and referees have a say in the final decision. The model adopted by the *Royal Society Open Science*, in turn, seems to strike the best balance between openness and privacy and could, in theory, be used as a template for political science journals. Their model allows for four possible scenarios: 1) if both author and referees agree to OPR, the review is made public 2) if only the referee agrees to OPR, his or her name is disclosed only to the author 3) if only the author agrees to OPR, the report is made public but the referee's name is not disclosed to the author or the public 4) if neither the author or referees agree to OPR, the referee report is not made public and the author does not know the referees' name.[10] Their method leaves room for all parts to reach an agreement about the publication of complementary texts and yet favors the open system.

If journals do not want to modify their current page layouts, one idea is to make their referee reports available on a data repository modelled after Dataverse or figshare.[11] This guarantees not only that reports are properly credited to the reviewers, but that any interested reader is able to access such reviews even without an institutional subscription. An online repository greatly simplifies the process of attributing a DOI number to any report at the point of publication and reviews can be easily shared and cited. In this model, journals would be free to choose between uploading the reports to an established repository or creating their own virtual services, akin to the several publication-exclusive dataverses that are already being used to store replication data. The same idea could be implemented by political science departments to illustrate the reviewing work of their faculty.

Since such large-scale changes may take some time to be achieved, referees can also publish their reports online individually if they believe in the merits of OPR. The startup Publons[12] offers a handy platform for reviewers to store and share their work, either as member of an institution or independently. Publons is free to use and can be linked to an ORCID profile with only few clicks.[13] For those concerned with privacy, reviewers retain total control over whatever they publish on Publons and no report is made public without explicit consent from all individuals involved in the publication process. More specifically, the referee must first ensure whether publicizing his or her review is in accordance with the journal policies and if the reviewed manuscript has already been either published in an academic journal and given a DOI. If these requirements are met, the report is made available online.

Publons also has a verification system for review records that allows referees to receive credit even if their full reports cannot be hosted on the website. For instance, reports for rejected manuscripts are not publicized in order to maintain the author's privacy,[14] and editorial policies often request reviewers to keep their reports anonymous. Verified review receipts circumvent these problems.[15] First, the scholar forwards his or her review receipts (i.e. e-mails sent by journal editors) to Publons. The website checks the authenticity of the reports, either automatically or by contacting editors, and notifies the referees that their reports have been successfully verified. At that moment, reviewers are free to make this information public.

Finally, Publons can be used as a post-publication platform for scholars, where they can comment on a manuscript that has already appeared in a journal. Post-publication reviews are yet uncommon in political science, but they may offer a good chance for PhD students to underscore their knowledge of a given topic and enter Publons' list of recommended reviewers. Since graduate students do not engage in traditional peer review as often as university professors, post-publication notes are one of the most effective ways for a junior researcher to join an ongoing academic debate.

Discussion

In this article I have tried to highlight that open peer review is an important yet overlooked tool to improve scientific exchange. It ensures higher levels of trust and reliability in academia, makes conflicts of interest explicit, lends credibility to referee reports, gives credit to reviewers, and allows others to scrutinize every step of the publishing process. While some scholars have stressed the possible drawbacks of the signed review system, open peer reviews rest on strong empirical and ethical grounds. Evidence from other fields suggests that signed reviews have better quality than their unsigned counterparts. At the very least, they promote research transparency without reducing the average quality of reports.

Scholars willing to share their reports online are encour-

[10]http://rsos.royalsocietypublishing.org/content/open-peer-review-royal-society-open-science/ (February 1, 2016)
[11]http://figshare.com/ (February 1, 2016)
[12]http://publons.com/ (February 1, 2016)
[13]https://orcid.org/blog/2015/10/12/publons-partners-orcid-give-more-credit-peer-review (February 1, 2016)
[14]https://publons.freshdesk.com/support/solutions/articles/5000538221-can-i-add-reviews-for-rejected-or-unpublished-manuscripts (February 1, 2016)
[15]https://publons.com/about/reviews/#reviewers (February 1, 2016)

aged to sign in to Publons and create an account. As I have also tried to show, there are several advantages for researchers who engage in the open system. As more scholars adopt signed reviewers, institutions may follow suit and support open peer reviews. The move towards more transparency in political science has increased the discipline's credibility to their own members and the public at large; open peer review is a further step in that direction.

Acknowledgments

I would like to thank Guilherme Duarte, Justin Esarey and David Skarbek for their comments and suggestions.

References

Bell, Mark S., and Nicholas L. Miller. 2015. "Questioning the Effect of Nuclear Weapons on Conflict." *Journal of Conflict Resolution* 59 (1): 74-92.

Broockman, David, Joshua Kalla, and Peter Aronow. 2015. "Irregularities in LaCour (2014)." May 19. `http://stanford.edu/~dbroock/broockman_kalla_aronow_lg_irregularities.pdf` (February 1, 2016).

DeCoursey, Tom. 2006. "Perspective: the Pros and Cons of Open Peer Review." *Nature*. DOI: 10.1038/nature04991.

Djupe, Paul A. 2015. "Peer Reviewing in Political Science: New Survey Results." *PS: Political Science & Politics* 48 (2): 346-52.

Eisenhart, Margaret. 2002. "The Paradox of Peer Review: Admitting Too Much or Allowing Too Little?" *Research in Science Education* 32 (2): 241-55.

Ernst, Edzard, and Thomas Kienbacher. 1991. "Correspondence: Chauvinism." *Nature* 352 (6336): 560.

Fabiato, Alexandre. 1994. "Anonymity of Reviewers." *Cardiovascular Research* 28 (8): 1134-9.

Fitzpatrick, Kathleen. 2011. *Planned Obsolescence: Publishing, Technology, and the Future of the Academy.* New York: New York University Press.

Ford, Emily. 2013. "Defining and Characterizing Open Peer Review: a Review of the Literature." *Journal of Scholarly Publishing* 44 (4): 311-26.

Ford, Emily. 2015. "Open Peer Review at Four STEM Journals: an Observational Overview." *F1000Research* 4 (6).

Gandrud, Christopher. 2013a. "GitHub: a Tool for Social Data Set Development and Verification in the Cloud." *The Political Methodologist* 20 (2): 7-16.

Gandrud, Christopher. 2013b. *Reproducible Research with R and R Studio.* CRC Press.

Godlee, Fiona. 2002. "Making Reviewers Visible: Openness, Accountability, and Credit." *Journal of the American Medical Association* 287 (21): 2762-5.

Greaves, Sarah, Joanna Scott, Maxine Clarke, Linda Miller, Timo Hannay, Annette Thomas, and Philip Campbell. 2006. "Overview: Nature's Peer Review Trial." *Nature*. DOI: 10.1038/nature05535.

Groves, Trish. 2010. "Is Open Peer Review the Fairest System? Yes." *BMJ* 341: c6424.

Hojat, Mohammadreza, Joseph S. Gonnella, and Addeane S. Caelleigh. 2003. "Impartial Judgment by the 'Gatekeepers' of Science: Fallibility and Accountability in the Peer Review Process." *Advances in Health Sciences Education* 8 (1): 75-96.

Jones, Zachary M. 2013. "Git/GitHub, Transparency, and Legitimacy in Quantitative Research." *The Political Methodologist* 21 (1): 6-7.

Khan, Karim. 2010. "Is Open Peer Review the Fairest System? No." *BMJ* 341: c6425.

King, Gary. 1995. "Replication, Replication." *PS: Political Science & Politics* 28 (3): 444-52.

King, Gary. 2006. "Publication, Publication." *PS: Political Science & Politics* 39 (1): 119-25.

Kowalczuk, Maria K., Frank Dudbridge, Shreeya Nanda, Stephanie L. Harriman, and Elizabeth C. Moylan. 2013. "A Comparison of the Quality of Reviewer Reports from Author-Suggested Reviewers and Editor-Suggested Reviewers in Journals Operating on Open or Closed Peer Review Models." *F1000 Posters* 4: 1252.

Kronick, David A. 1990. "Peer Review in 18th-Century Scientific Journalism." *Journal of the American Medical Association* 263 (10): 1321-2.

Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. 2013. "Bias in Peer Review." *Journal of the American Society for Information Science and Technology* 64 (1): 2-17.

Lipworth, Wendy, Ian H. Kerridge, Stacy M. Carter, and Miles Little. 2011. "Should Biomedical Publishing Be "Opened up"? Toward a Values-Based Peer-Review Process." *Journal of Bioethical Inquiry* 8 (3): 267-80.

Lupia, Arthur, and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency." *PS: Political Science & Politics* 47 (1): 19-42.

McNutt, Marcia. 2015. "Editorial Retraction." *Science* 348 (6239): 1100.

McNutt, Robert A., Arthur T. Evans, Robert H. Fletcher, and Suzanne W. Fletcher. 1990. "The Effects of Blinding on the Quality of Peer Review: a Randomized Trial." *Journal of the American Medical Association* 263 (10): 1371-6.

Meadows, Arthur Jack. 1998. *Communicating Research.* San Diego: Academic Press.

Peters, Douglas P., and Stephen J. Ceci. 1982. "Peer-Review Practices of Psychological Journals: the Fate of

Published Articles, Submitted Again." *Behavioral and Brain Sciences* 5 (2): 187-95.

Pöschl, Ulrich. 2012. "Multi-Stage Open Peer Review: Scientific Evaluation Integrating the Strengths of Traditional Peer Review with the Virtues of Transparency and Self-Regulation." *Frontiers in Computational Neuroscience* 6: 33.

Rennie, Drummond. 1998. "Freedom and Responsibility in Medical Publication: Setting the Balance Right." *Journal of the American Medical Association* 280 (3): 300-2.

Ross, Joseph S., Cary P. Gross, Mayur M. Desai, Yuling Hong, Augustus O. Grant, Stephen R. Daniels, Vladimir C. Hachinski, Raymond J. Gibbons, Timothy J. Gardner, and Harlan M. Krumholz. 2006. "Effect of Blinded Peer Review on Abstract Acceptance." *Journal of the American Medical Association* 295 (14): 1675-80.

Rowland, Fytton. 2002. "The Peer-Review Process." *Learned Publishing* 15 (4): 247-58.

Sandström, Ulf, and Martin Hällsten. 2008. "Persistent Nepotism in Peer-Review." *Scientometrics* 74 (2): 175-89.

Shanahan, Daniel R., and Bjorn R. Olsen. 2014. "Opening Peer-Review: the Democracy of Science." *Journal of Negative Results in Biomedicine* 13: 2.

Smith, Richard. 2006. "Peer Review: a Flawed Process at the Heart of Science and Journals." *Journal of the Royal Society of Medicine* 99 (4): 178-82.

Spier, Ray. 2002. "The History of the Peer-Review Process." *TRENDS in Biotechnology* 20 (8): 357-8.

Van Rooyen, Susan, Fiona Godlee, Stephen Evans, Nick Black, and Richard Smith. 1999. "Effect of Open Peer Review on Quality of Reviews and on Reviewers' Recommendations: a Randomised Trial." *BMJ* 318 (7175): 23-7.

Van Rooyen, Susan, Tony Delamothe, Stephen J. W. Evans. 2010. "Effect on Peer Review of Telling Reviewers That Their Signed Reviews Might Be Posted on the Web: Randomised Controlled Trial." *BMJ* 341: c5729.

Walsh, Elizabeth, Maeve Rooney, Louis Appleby, and Greg Wilkinson. 2000. "Open Peer Review: a Randomised Controlled Trial." *The British Journal of Psychiatry* 176 (1): 47-51.

Ware, Mark. 2008. "Peer Review in Scholarly Journals: Perspective of the Scholarly Community-Results from an International Study." *Information Services and Use* 28 (2): 109-12.

Weller, Ann C. 2001. *Editorial Peer Review: Its Strengths and Weaknesses*. Medford, New Jersey: Information Today, Inc.

Wendler, David, and Franklin Miller. 2014. "The Ethics of Peer Review in Bioethics." *Journal of Medical Ethics* 40 (10): 697-701.

Wennerås, Christine, and Agnes Wold. 1997. "Nepotism and Sexism in Peer Review." *Nature* 387 (6631): 341-3.

The Multiple Routes to Credibility

Thomas J. Leeper
London School of Economics and Political Science
thosjleeper@gmail.com

Peer review – by which I mean the decision to publish research manuscripts based upon the anonymous critique of approximately three individuals – is a remarkably recent innovation. As a June, 2015 article in *Times Higher Education* makes clear, the content of scholarly journals has been historically decided almost entirely by the non-anonymous read of one known individual, namely the journal's editor (Fyfe 2015). Indeed, even in the earliest "peer-reviewed" journal, *Philosophical Transactions*, review was conducted by something much closer to an editorial board than a broad, blind panel of scientific peers (Spier 2002, 357). Biagioli (2002) describes how early academic societies in 17th century England and France implemented review-like procedures less out of concern for scientific integrity than out of concern for their public credibility. Publishing questionable or objectionable content under the banner of their society was risky; publishing only papers from known members of the society or from authors who had been vetted by such members became policy early on (Bornmann 2013).

Despite these early peer reviewing institutions, major scientific journals retained complete discretionary editorial control well into the 20th century. Open science advocate Michael Nielsen highlights that there is only documentary evidence that one of Einstein's papers was ever subject to modern-style peer review (Nielsen 2009). The journal *Nature* adopted a formal system of peer review only in 1967. *Public Opinion Quarterly*, as another example, initiated a peer review process only out of necessity in response to "impressive numbers of articles from contributors who were outside the old network and whose capabilities were not known to the editor" (Davison 1987, S8) and even then largely involved members of the editorial board.[1]

While scholars are quick to differentiate a study "not yet peer reviewed" from one that has passed this invisible

[1] Campanario (1998a) and Campanario (1998b) provides an impressively comprehensive overview of the history and present state of peer review as an institution.

threshold, peer review is but one route to scientific credibility and not a particularly reliable method of enhancing scientific quality given its present institutionalization. In this article, I argue that peer review must be seen as part of a set of scientific practices largely intended to enhance the credibility of truth claims. As disciplines from political science to psychology to biomedicine to physics debate the reproducibility, transparency, replicability, and truth of their research literatures,[2] we must find a proper place for peer review among a host of individual and disciplinary activities. Peer review should not be abandoned, but it must be reformed in order to retain value for contemporary science, particularly if it is also meant to improve the quality of science rather than simply the perception that scientific claims are credible.

The Multiple Routes to Credibility

As Skip Lupia and Colin Elman have argued, anyone in society can make knowledge claims (Lupia and Elman 2014). The sciences have no inherent privilege nor uncontested authority in this regard. Science, however, distinguishes itself through multiple forms of claimed credibility. Peer review is one such path to credibility. When mainstream scientists make claims, they do so with the provisional non-rejection of (hopefully) similarly or more able peers. Regardless of what peer review actually does to the content or quality of science,[3] it collectivizes an otherwise individual activity, making claims appear more credible because once peer reviewed they carry the weight of the scientific societies orchestrating the review process.

Peer review as currently performed is essentially an outward-facing activity. It enhances the standing of scientific claims relative to claims made by others above and beyond the credibility lent simply by one being a scientist as opposed to someone else (Hovland and Weiss 1951; Pornpitakpan 2004). As such, peer review is not an essential aspect of science, but rather a valuable aspect of science communication. Were we principally concerned with the impact of peer review on scientific quality (as opposed to the appearance of scientific quality), we would (1) rightly acknowledge the informal peer review that almost all research is subject to (through presentation, informal conversation, and public discussion), (2) conduct peer review earlier and perhaps multiple times during the execution of the scientific method (rather than simply at is conclusion), and (3) subject all forms of public scientific claim-making (such as conference presentations, book publication, blog and social media posts, etc., which are only reviewed in some cases) to more rigorous, centralized peer review. These things we do not do because – as has always been the case – peer review is primarily about the protection of the credibility of the groups publishing research.[4] If peer review were the only way of enhancing the credibility of knowledge claims, this slightly impotent system of post-analysis/pre-publication review might make sense, particularly if it had demonstrated value apart from its contribution to scientific credibility.

Yet, the capacity of peer review processes to enhance rather than diminish scientific quality is questionable, given how peer review introduces well-substantiated publication biases (Sterling 1959; Sterling, Rosenbaum, and Weinkam 1995; Gelman and Weakliem 2009), is highly unreliable,[5] and offers a meager barrier to scientific fraud (as political scientists know all too well).

Alternative yet complementary paths to credibility that are unique to science include the three other R's: registration, reproducibility, and replication. *Registration* is the documentation of scientific projects prior to being conducted (Humphreys, Sanchez de la Sierra, and van der Windt 2013). Registration aims to avoid publication biases, in particular the "file drawer" problem (Rosenthal 1979). Registering and ideally pre-accepting such kernels of research provides a uniquely powerful way of combating publication biases. Registration indicates that research was worth doing regardless of what it finds and offers some partial guarantee that the results were not selected for their size or substance. The real advantage of registration will come when it is combined with processes of pre-implementation peer review (which I argue for below) because publications will be based on the degree to which they are interesting, novel, and valid without regard to the particular findings that result from that research. Pilot tests of such

[2] See both now 25-year-old debate in *The Journal of the American Medical Association* reporting on the First International Congress in Peer Review in Biomedical Publication (Lundberg 1990) and a recent debate in *Nature* (Nature's peer review debate 2006) for an excellent set of perspectives on peer review.

[3] Though only one meta-analysis, a 2007 *Cochrane Collaboration* systematic review of studies of peer review found no evidence of impact of peer review on quality of biomedical publications (Jefferson et al. 2007). More anecdotally, Richard Smith, long-time editor of the *The British Medical Journal* has repeatedly criticized peer review as irrelevant (Smith 2006) even to the point of arguing for its abolition (Smith 2015).

[4] Justin Esarey rightly points out that "peer review" is used here in the sense of a formal institution. One could argue that there is – and has always been – a peer reviewing market at work that comments on, critiques, and responds to scientific contributions. This is most visible in social media discussions of newly released research, but also in the discussant comments offered at conferences, and in the informal conversations had between scholars. Returning to the Einstein example from earlier, the idea of general relativity was never peer reviewed prior to or as part of publication but has certainly been subjected to review by generations of physicists.

[5] Cicchetti (1991) provides a useful meta-analysis of reliability of grant reviews. Peters and Ceci (1982), in a classic study, found highly cited papers once resubmitted were frequently rejected by the same journals that originally published them with little recognition of the plagiarism. For a more recent study of consistency across reviews, see this blog post about the 2014 NIPS conference (Price 2014). And, again, see Campanario (1998a, 1998b) for a thorough overview of the relevant literature.

pre-registration reviewing at *Comparative Political Studies*, *Cortex*, and *The Journal of Experimental Political Science* should prove quite interesting.

Reproducibility relates to the transparency of conducted research and the extent to which data sources and analysis are publicly documented, accessible, and reusable (Stodden, Guo, and Ma 2013). Reproducible research translates a well-defined set of inputs (e.g., raw data sources, code, computational environment) into the set of outputs used to make knowledge claims: statistics, graphics, tables, presentations, and articles. It especially avoids consequential errors that result in (un)intended misreporting of results (see, provocatively, Nuijten et al. N.d). Reproducibility invites an audience to examine results for themselves, offering self-imposed public accountability as a route to credibility.

Replication involves the repeated collection and analysis of data along a common theme. Among the three R's, replication is clearly political science's most wanting characteristic. Replication – due to its oft-claimed lack of novelty – is seen as secondary science, unworthy of publication in the discipline's top scientific outlets. Yet replication is what ensures that findings are not simply the result of sampling errors, extremely narrow scope conditions, poorly implemented research protocols, or some other limiting factor. Without replication, all we have is Daryl Bem's claims of precognition (French 2012). Replication builds scientific literatures that offer far more than a collection of ad-hoc, single-study claims.[6] Replication further invites systematic review that in turn documents heterogeneity in effect sizes and the sensitivity of results to samples, settings, treatments, and outcomes (Shadish, Cook, and Campbell 2001).

These three R's of modern, open science are complementary routes to scientific credibility alongside scientific review. The challenge is that none of them guarantees scientific quality. Registered research might be poorly conceived, reproducible research might be laden with errors, and replicated research might be pointless or fundamentally flawed. Review, however, is uniquely empowered to improve the quality of science because of its capacity for both binding and conversational input from others. Yet privileging peer review over these other forms of credibility enhancement prioritizes an anachronistic, intransparent, and highly contrived social interaction of ambiguous effect over alternatives with face-valid positive value. The three R's are credibility enhancing because they offer various forms of lasting public accountability. Peer review, by contrast, does not. How then can review be improved in order to more transparently enhance quality and thus lend further credibility to scientific claims? The answer comes in both institutional reforms and changes in reviewer behavior.

Toward Better Peer Review

First, several institutional changes are clearly in order:

1. **Greater transparency**. The peer review process produces an enormous amount of metadata, which should all be public (including reviews themselves). Without such information, it is difficult to evaluate the effectiveness of the institution and without transparency, there is little opportunity for accountability in the process. Freire (2015) makes a compelling argument for how this might work in political science. At a minimum, releasing data on author characteristics, reviewer characteristics and decisions, and manuscript topics (and perhaps coding of research contexts, methods, and findings) would enable exploratory research on correlates of publication decisions. Releasing reviews themselves after an embargo period, would hold reviewers (anonymously) accountable for the content and quality of reviews and enable further research into, perhaps subtle, biases in the reviewing process. Accountability for the content of reviews and for peer review decisions should – following the effects of accountability generally (Lerner and Tetlock 1999) – improve decision quality.

2. **Earlier peer review**. Post-analysis peer review is an immensely ineffective method of enhancing scientific quality. Reviewers are essentially fettered, able only to comment on research that has little to no hope of being conducted anew. This invites outcome-focused review and superficial, editorial commentary. Peer review should instead come earlier in the scientific process, ideally prior to analysis and prior to data collection, when it would still be possible to change the theories, hypotheses, data collection, and planned analysis. (And when it might filter out research that is so deficient that resources should not be expended on it.) If peer review is meant to affect scientific quality, then it must occur when it has the capacity to actually affect science rather than manuscripts. Such review would replace post-analysis publication and would need to be binding on journals, so that they cannot refuse to publish "uninteresting" results of otherwise "interesting" studies. Pre-reviewed research should also have higher credibility because it has a plausible claim of objectivity: reporting is not based on novelty or selective reporting of ideologically favorable results.[7] If peer

[6] It is important to caution, however, against demanding multi-study papers. Conditioning publication of a given article on having replications introduces a needless file drawer problem with limited benefit to scientific literature. Schimmack (2012), for example, shows that multi-study papers appear more credible than single-study papers even when the latter approach has higher statistical power.

[7] Some opportunities for pre-analysis peer review already exist, including: Time-Sharing Experiments for the Social Sciences, funding agency review processes, and some journal efforts (such as in a recent *Comparative Political Studies* special issue or the new "registration" track of the *Journal of Experimental Political Science*).

review continues to occur after data are collected and analyzed, a lesser form of "outcome-blind" reviewing could at least constrain reviewer-induced publication biases.

3. **Bifurcated peer review**. Journals, owned by for-profit publishers and academic societies, care about rankings. It is uncontroversial that this invites a focus on research outcomes sometimes at the expense of quality. It also means that manuscripts often cycle through numerous review processes before being published, haphazardly placing the manuscript's fate in the hands of a sequence of three individuals. Removing peer review from the control of journals would streamline this process, subjecting a manuscript to only one peer review process, the conclusion of which would be a competitive search for the journal of best fit.[8] Because review is not connected to any specific outlet, reviews can focus on improving quality without considering whether research is "enough" for a given journal. A counterargument would be that bifurcation means centralization, and with it an even greater dependence on unreliable reviews. To the contrary, a centralized process would be better able to equitably distribute reviewer workloads, increase the number of reviews per manuscript,[9], potentially increase reviewers' engagement with and commitment to improving a given manuscript, and enable the review process to involve a dialogue between authors and reviewers rather than a one-off, one-way communication. Like a criminal trial, one could also imagine various experimental innovations such as allowing pre-review stages of evidence discovery and reviewer selection. But the overall goal of a bifurcated process would be to publish more research, publish concerns about that research alongside the manuscripts, and allow journals to focus on recommending already reviewed research.

4. **Post-publication peer review**. Publicly commenting on already published manuscripts through venues like F1000Research, The Winnower, or RIO helps to put concerns, questions, uncertainties, and calls for replication and future research into the permanent record of science. Research is constantly "peer reviewed" through the normal process of scientific discussion and occasional replication, but discussions are rarely recorded as expressed concerns about manuscripts, so post-publication peer review ensures that publication is not the final say on a piece of research.

While these reforms are attractive ways to improve the peer review process, widespread reform is probably long off. How can reviewers behave today in order to enhance the credibility of scientific research while also (ideally) improving the actual quality of scientific research? Several ideas come to mind:

1. **Avoid introducing publication bias**. As a scientist, the size, direction, and significance of a finding should not affect whether I see that research as well-conducted or meriting wider dissemination.[10] This requires outcome-blind reviewing, even if self-imposed. It also means that novelty is not a useful criterion when evaluating a piece of research.

2. **Focus on the science, not the manuscript**. An author has submitted a piece of research that they feel moderately confident about. It is not the task of reviewers to rewrite the manuscript; that is the work of the authors and editors. Reviewers need to focus on science, not superficial issues in the manuscript. A reviewer does not have to be happy with a manuscript, they simply have to not be dissatisfied with the apparent integrity of the underlying science.

3. **Consider reviewing a conversation**. The purpose is to enhance scientific quality (to the extent possible after research has already been conducted). This means that lack of clarity should be addressed through questions not rejections nor demands for alternative theories. Reviews should not be shouted through the keyboard but rather seen as the initiation of a dialogue intended to equally clarify both the manuscript and the reviewer's understanding of the research. In short, be nice.

4. **Focus on the three other R's**. Reviewers should ensure that authors have engaged in transparent, reproducible reporting and they should reward authors engaged in registration, reproducibility, and replication. If necessary, they should ask to see the data or

[8] Or, ideally, the entire abandonment of journals as an antiquated, space-restricted venue for research dissemination.

[9] Overall reviewer burden would be constant or even decrease because the same manuscript would not be required to be passed to a completely new set of reviewers at each journal. Given that current reviewing process involve a small number of reviewers per manuscript, the results of review processes tend to have low reliability (Bornmann, Mutz, and Daniel 2010); increasing the number of reviewers would tend to counteract that.

[10] Novel, unexpected, or large results may be exciting and publishing them may enhance the impact factor or standing of a journal. But if peer review is to focus on the core task of enhancing scientific quality, then those concerns are irrelevant. If peer reviewer were instead designed to identify the most exciting, most "publishable" research, then authors should be monetarily incentivized to produce such work and peer reviewers should be paid for their time in identifying such studies. The public credibility of claims published in such a journal would obviously be diminished, however, highlighting the need for peer review as an exercise in quality alone as a route to credibility. One might argue that offering such judgments are helpful because editors work with a finite number of printable pages. Such arguments are increasingly dated, as journals like *PLoS* demonstrate that there is no upper bound to scientific output once the constraints of print publication are removed.

features thereof to address concerns, even those materials cannot be made a public element of the research. They should never demand analytic fishing expeditions, quests for significance and novelty, or post-hoc explanations for features of data.

These are simple principles, but it is surprising how rarely they are followed. To maximize its value, review must be much more focused on scientific quality than novelty-seeking and editorial superficiality. The value of peer review is that it can actually improve the quality of science, which in turn makes those claims more publicly credible. But it can also damage scientific quality. This means that the activity of peer review within current institutional structures should take a different tone and focus, but the institution itself should also be substantially reformed. Further, political science should be in the vanguard of adopting practices of registration, reproducibility, and replication to broadly enhance the credibility of our collective scientific contribution. Better peer review will ensure those scientific claims not only appear credible but actually are credible.

References

Biagioli, Mario. 2002. "From Book Censorship to Academic Peer Review." *Emergences: Journal for the Study of Media & Composite Cultures* 12 (1): 11-45.

Bornmann, Lutz. 2013. "Evaluations by Peer Review in Science." *Springer Science Reviews* 1 (1): 1-4.

Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2010. "A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants." *PLoS ONE* 5 (12): e14331.

Campanario, Juan Miguel. 1998a. "Peer Review for Journals as It Stands Today – Part 1." *Science Communication* 19 (3): 181-211.

Campanario, Juan Miguel. 1998b. "Peer Review for Journals as It Stands Today – Part 2." *Science Communication* 19 (4): 277-306.

Cicchetti, Domenic V. 1991. "The Reliability of Peer Review for Manuscript and Grant Submissions: A Cross-Disciplinary Investigation." *Behavioral and Brain Sciences* 14 (1): 119-35.

Davison, W. Phillips. 1987. "A Story of the POQ's Fifty-Year Odyssey." *Public Opinion Quarterly* 51: S4-S11.

Freire, Danilo. 2015. "Peering at Open Peer Review." *The Political Methodologist* 23 (1): 6-11.

French, Chris. 2012. "Precognition Studies and the Curse of the Failed Replications." *The Guardian*, March 15. `https://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications` (February 1, 2016).

Fyfe, Aileen. 2015. "Peer Review: Not as Old As You Might Think." June 25, 2015. `https://www.timeshighereducation.com/features/peer-review-not-old-you-might-think` (February 1, 2016).

Gelman, Andrew, and David Weakliem. 2009. "Of Beauty, Sex and Power." *American Scientist* 97 (4): 310.

Hovland, Carl I., and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15 (4): 635-50.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21 (1): 1-20.

Jefferson, Tom, Melanie Rudin, Suzanne Brodney Folse, and Frank Davidoff. 2007. "Editorial Peer Review for Improving the Quality of Reports of Biomedical Studies." *Cochrane Database of Systematic Reviews* 2: MR000016.

Lerner, Jennifer S., and Philip E. Tetlock. 1999. "Accounting for the Effects of Accountability." *Psychological Bulletin* 125 (2): 255-75.

Lundberg, George D. (Ed.) 1990. *The Journal of the American Medical Association* 263 (10).

Lupia, Arthur, and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency." *PS: Political Science & Politics* 47 (1): 19-42.

Nature's peer review debate. 2006. *Nature*. `http://www.nature.com/nature/peerreview/debate/index.html` (February 1, 2016)

Nielsen, Michael. 2009. "Three Myths About Scientific Peer Review." January 8, 2009. `http://michaelnielsen.org/blog/three-myths-about-scientific-peer-review` (February 1, 2016)

Nuijten, Michèle B., Chris H. J. Hartgerink, Marcel A. L. M. van Assen, Sacha Epskamp, and Jelte M. Wicherts. N.d. "The Prevalence of Statistical Reporting Errors in Psychology (1985-2013)." *Behavior Research Methods*. Forthcoming.

Peters, Douglas P., and Stephen J. Ceci. 1982. "Peer-Review Practices of Psychological Journals: the Fate of Published Articles, Submitted Again." *Behavioral and Brain Sciences* 5 (2): 187-95.

Pornpitakpan, Chanthika. 2004. "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34 (2): 243-81.

Price, Eric. 2014. "The NIPS Experiment." December 15. `http://blog.mrtz.org/2014/12/15/the-nips-experiment` (February 1, 2016).

Rosenthal, Robert. 1979. "The "File and Drawer Problem" and Tolerance for Null Results." *Psychological Bulletin* 86 (3): 638-41.

Schimmack, Ulrich. 2012. "The Ironic Effect of Significant Results on the Credibility of Multiple-Study Articles." *Psychological Methods* 17(4): 551-66.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company.

Smith, Richard. 2006. "Peer Review: a Flawed Process at the Heart of Science and Journals." *Journal of the Royal Society of Medicine* 99 (4): 178-82.

Smith, Richard. 2015. "The Peer Review Drugs Don't Work." May 28. https://www.timeshighereducation.com/content/the-peer-review-drugs-dont-work (February 1, 2016).

Spier, Ray. 2002. "The History of the Peer-Review Process." *TRENDS in Biotechnology* 20 (8): 357-8.

Sterling, Theodore D. 1959. "Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance – Or Vice Versa." *Journal of the American Statistical Association* 54 (285): 30-4.

Sterling, Theodore D., Wilf L. Rosenbaum, and James J. Weinkam. 1995. "Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa." *The American Statistician* 49 (1): 108-12.

Stodden, Victoria, Peixuan Guo, and Zhaokun Ma. 2013. "Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals." *PLoS ONE* 8 (6): e67111.

What is Peer Review For? Why Referees are not the Disciplinary Police

Thomas Pepinsky
Cornell University
pepinsky@cornell.edu

The peculiar thing about peer review is that it is central to our professional lives as political scientists, yet we tend to talk about refereeing only in the most general and anonymous terms. I have never shown an anonymous referee report on my own submitted manuscripts to anyone, ever. The only other times I have seen other referee reports is when a journal gives me access to other reports for a manuscript that I have refereed. There is very little scope for dialogue on the value of particular referee reports; indeed, as an institution, peer review precludes dialogue between authors and referees unless, of course, the manuscript has already made it past the first stage! That means there is almost certainly a great diversity of viewpoints within the profession about what makes a good referee report, or about what exactly a referee report is supposed to do. Miller et al. (2013), discuss the components of a good referee report, but leave unstated the function of the review itself, aside from to note that it is important for the scholarly process.

In my own view, a referee report has three functions. It is first a *recommendation to an editor*. This is the literal function of peer review, to communicate to the journal about whether or not the manuscript should be published, and under what conditions. The second function of a referee report is to provide *comments to the author*. The act of evaluating a manuscript entails explaining what its strengths and weaknesses are, and those evaluations are not just for the editors, they are comments for the author as well. Some referees also provide more than just comments on what is working and what is not, making suggestions about how the manuscript might be improved.

The third function of a referee report is an indirect one: *shaping the discipline*. Because editors use manuscript evaluations to decide whether or not to publish submissions, it follows that referee reports affect what gets published. Because publication is – appropriately – so central to the discipline, and because top journal slots are scarce, referee reports inevitably determine the direction of political science research.

These three functions of peer review differ from one another, and it follows that every referee report is necessarily doing many things at the same time. I suspect that the institution of peer review is probably not ideally suited for doing any of them. For example, the anonymity of peer review enables a kind of forthrightness that might not otherwise be possible with a personal interaction, but the impossibility of dialogue (unless the manuscript already has a good chance of being published) makes little sense if the goal is to provide meaningful feedback to authors. Peer review does effectively shape the discipline, but it may do so in inefficient and even counterproductive ways by discouraging novel, critical, or outside-the-box thinking.

Among political scientists, there are surely very different ideas about which of these three peer review functions are the important ones, and it is here where the anonymity of peer review as an institution makes it difficult to establish common expectations about what makes a referee report good or bad, or useful or not. My own tastes may be peculiar, but I attempt to focus exclusively on recommendations to the editor and comments to the author in my own referee reports. That is, I strive to be indifferent to concerns of the type *if this manuscript is published, then people will work on this topic or adopt this methodology, even if I think it is* *boring or misleading?* Instead, I try to focus on questions like *is this manuscript accomplishing what it sets out to accomplish?* and *are there ways to my comments can make it better?* My goal is to judge the manuscript on its own terms.

Why? Because as scientists, we must be radically skeptical that there is a single model for the discipline, or that we can identify *ex ante* what makes a contribution valuable. That is a case for the author to make. If we accept this, then true purpose of the referee report is simply to evaluate whether or not that case has been made.

Readers will note that focusing on feedback to the author and recommendations to the editor discourages certain kinds of commentary when evaluating a manuscript. The comment that *this journal should not publish this kind of work*, for example, is not relevant to my evaluation. I consider this to be a judgment for the editors, and I assume (perhaps naively) that any manuscript that I have been asked to referee has not been ruled out on grounds of its theoretical or methodological approach. *I do not know how to evaluate this kind of research* is also not a useful comment for either the editor or the author.

Most importantly, the comment that *studies of this type inherently suffer from a fatal flaw* only makes sense as a way to provide feedback about how a manuscript might be improved, not as a summary judgment against the manuscript. Imagine, for example, a referee who is implacably opposed to all survey experiments. My views about the function of peer review would hold that his or her referee report should provide concrete feedback about how the manuscript could be improved, with specific attention to whatever proposed flaw stems from the use of survey experiments. This view, in other words, does not insist that referees accept methodologies or frameworks that they find to be problematic. It does demand that referees be able to articulate directions for improvement; in this example, ways to complement the inferences that can be drawn from survey experiments with other forms of data, or suggestions about how the author might acknowledge limitations.

My views about what referee reports are for have evolved over time. Many younger scholars probably believe, as I once did, that the function of the referee is to act the part of the disciplinary police officer, to "protect" the community from "bad" research. What I now believe is that it is much harder to identity objectively "bad" research than we think, and the best way to orient my referee reports is around the question identified above: *does this manuscript accomplish what the author sets out to do?* These changing tastes on my part are probably a result of my experiences as an author, because I hate reading referee reports that conclude that my manuscripts should be about something else. And so I try to referee manuscripts with this concern in mind, and am especially mindful of this for manuscripts that do not match my own tastes.

There are two additional benefit of approaching peer review this way. First, it encourages *the referee* to try to articulate what the goal of the manuscript is. This should be obvious from the submission itself, and if it is not, then that is one big strike against the manuscript. (One useful exercise when writing a referee report is simply to summarize the manuscript.) Second, it should work against the incentives for disciplinary narrowness that are inherent to peer review. Some manuscripts are narrow and precise. Some manuscripts strive to be provocative. Some manuscripts seek to integrate disparate theories or literatures. Working from the assumption that each of these types of submissions can be valuable, judging a manuscript with an eye towards the author's aims should encourage more creative research without punishing research in the normal science vein.

These points notwithstanding, it is surely true that referee tastes affect their evaluations, and to pretend otherwise is counterproductive. I suspect that editors have massive influence over the fate of individual manuscripts in choosing referees, and also in weighing different referees' evaluations when their evaluations diverge.

Editorial judgments are probably most important when considering cross-disciplinary work, for the functions of peer review almost certainly vary across disciplines. I have refereed manuscripts from anthropology, religious studies, and Asian studies, among others. Before agreeing to referee these manuscripts, I always make clear to editors that they are going to get a referee report that follows my understanding of the conventions of political science. My logic is "if you go to a barber, you'll get a haircut," so I want to make sure that editors know that I am a barber, and so they should not expect a sandwich.

These presumed differences aside, no editor has ever thought that my training as a political scientist disqualifies me; conditional on inviting me to referee a manuscript, editors uniformly pledge accept my disciplinary biases. Even so, referees have a particular responsibility to authors of manuscripts outside of their disciplinary traditions, not because our evaluations are necessarily negative (I have indeed recommended in favor of publication in some cases), but because they probably draw on an entirely different understanding of what makes research valuable, and how authors can demonstrate that their work has accomplished the goals that they set out for themselves. Without such a shared understanding, it is hard to see how we are "peers" in any meaningful sense.

This point brings us back to my initial observation, that as a discipline we talk about referee reports only in the most general and anonymous terms, without consensus about what peer review is for. It helps, when receiving a painful negative review, to recognize that referees may have very different assumptions about what their reports are supposed to do. A more constructive focus on editorial recommendations and author comments will never soothe the sting of a rejection (or a "decline"; see Isaac 2015), but it could ensure that the dialogue remains focused on what the manuscript hopes to accomplish. And that – not policing the discipline – is what peer review is for.

References

Isaac, Jeffrey C. 2015. "Beyond "Rejection."" *Duck of Minerva* December 2. `http://duckofminerva.com/2015/12/beyond-rejection.html` (February 1, 2016).

Miller, Beth, Jon Pevehouse, Ron Rogowski, Dustin Tingley, and Rick Wilson. 2013. "How To Be a Peer Reviewer: A Guide for Recent and Soon-to-be PhDs." *PS: Political Science & Politics* 46 (1): 120-3.

An Editor's Thoughts on the Peer Review Process

Sara McLaughlin Mitchell
University of Iowa
sara-mitchell@uiowa.edu

As academics, the peer review process can be one of the most rewarding and frustrating experiences in our careers. Detailed and careful reviews of our work can significantly improve the quality of our published research and identify new avenues for future research. Negative reviews of our work, while also helpful in terms of identifying weaknesses in our research, can be devastating to our egos and our mental health. My perspectives on peer review have been shaped by twenty years of experience submitting my work to journals and book publishers and by serving as an Associate Editor for two journals, *Foreign Policy Analysis* and *Research & Politics*. In this piece, I will 1) discuss the qualities of good reviews, 2) provide advice for how to improve the chances for publication in the peer review process, and 3) discuss some systemic issues that our discipline faces for ensuring high quality peer review.

Let me begin by arguing that we need to train scholars to write quality peer reviews. When I teach upper level graduate seminars, I have students submit a draft of their research paper about one month before the class ends. I then assign two other students as peer reviewers for the papers anonymously and then serve as the third reviewer myself for each paper. I send students examples of reviews I have written for journals and provide general guidelines about what improves the qualities of peer reviews. After distributing the three peer reviews to my students, they have two weeks to revise their papers and write a memo describing their revisions. Their final research paper grade is based on the quality of their work at each of these stages, including their efforts to review classmates' research projects.[1]

Writing High Quality Peer Reviews

What qualities do good peer reviews share? My first observation is that tone is essential to helping an author improve their research. If you make statements such as "this was clearly written by a graduate student" or "this paper is not important enough to be published in journal X" or "this person knows nothing about the literature on this topic", you are not being helpful. These kinds of blanket negative statements can only serve to discourage junior (and senior!) scholars from submitting work to peer reviewed outlets.[2] Thus one should always consider what contributions a paper is making to the discipline and then proceed with ideas for making the final product better.

In addition to crafting reviews with a positive tone, I also recommend that reviewers focus on criticisms internal to the project rather than moving to a purely external critique. For example, suppose an author was writing a paper on the systemic democratic peace. An internal critique might point to other systemic research in international relations that would help improve the authors' theory or identify alternative ways to measure systemic democracy. An external critique, however, might argue that systemic research is not useful for understanding the democratic peace and that the author should abandon this perspective in favor of dyadic analyses. If you find yourself writing reviews where you are telling authors to dramatically change their research question or theoretical perspective, you are not helping them produce publishable research. As an editor, it is much more helpful to have reviews that accept the authors' research goals and then provide suggestions for improvement. Along these lines, it is very common for reviewers to say things like "this person does not know the literature on the democratic peace" and then fail to provide a single citation for research

[1] While I am fairly generous in my grading of students' peer reviews given their lack of experience, I find that I am able to discriminate in the grading process. Some students more effectively demonstrate that they read the paper carefully, offering very concrete and useful suggestions for improvement. Students with lower grades tend to be those who are reluctant to criticize their peers. Even though I make the review process double blind, PhD students in my department tend to reveal themselves as reviewers of each other's work in the middle of the semester.

[2] In a nice piece that provides advice on how to be a peer reviewer, Miller et al. (2013, 122) make a similar point: "There may be a place in life for snide comments; a review of a manuscript is definitely not it."

[3] As Miller et al. (2013, 122) note: "Broad generalizations – for instance, claiming an experimental research design 'has no external validity' or merely stating 'the literature review is incomplete' – are unhelpful."

that is missing in the bibliography.[3] If you think an author is not engaging with an important literature for their topic, help the scholar by citing some of that work in your review. If you do not have time to add full citations, even providing authors' last names and the years of publication can be helpful.

Another common strategy that reviewers take is to ask for additional analyses or robustness checks, something I find very useful as a reader of scholarly work. However, reviewers should identify new analyses or data that are essential for checking the robustness of the particular relationship being tested, rather than worrying about all other control variables out there in the literature or all alternative statistical estimation techniques for a particular problem. A person reviewing a paper on the systemic democratic peace could reasonably ask for alternative democracy measures or control variables for other major systemic dynamics (e.g. World Wars, hegemonic power). Asking the scholar to develop a new measure for democracy or to test her model against all other major international relations systemic theories is less reasonable. I understand the importance for checking the robustness of empirical relationships, but I also think we can press this too far when we expect an author to conduct dozens of additional models to demonstrate their findings. In fact, authors are anticipating that reviewers will ask for such things and they are preemptively responding by including appendices with additional models. In conversations with my junior colleagues (who love appendices!), I have noted that they are doing a lot of extra work on the front end and getting potentially fewer publications from these materials when they relegate so much of their work to appendices. Had Bruce Russett and John Oneal adopted this strategy, they would have published one paper on the Kantian tripod for peace, rather than multiple papers that focused on different legs of the tripod. I also feel that really long appendices are placing additional burdens on reviewers who are already paying costs to read a 30+ page paper.[4]

Converting R&Rs to Publications

Once an author receives an invitation from a journal to revise and resubmit (R&R) a research paper, what strategies can they take to improve their chances for successfully converting the R&R to a publication? My first recommendation is to go through each review and the editors' decision letter and identify each point being raised. I typically move each point into a document that will become the memo describing my revisions and then proceed to work on the revisions. My memos have a general section at the beginning that provides an overview of the major revisions I have undertaken and then this is followed by separate sections for the editors' letter and each of the reviews. Each point that is addressed by the editors or reviewers is presented and then I follow that with information about how I revised the paper in light of that comment and the page number where the revised text or results can be found. It is a good idea to identify criticisms that are raised by multiple reviewers because these issues will be very imperative to address in your revisions. You should also read the editors' letter carefully because they often provide ideas about which criticisms are most important to address from their perspective. Additional robustness checks that you have conducted can be included in an appendix that will be submitted with the memo and your revised paper.

As an associate editor, I have observed authors failing at this stage of the peer review process. One mistake I often see is for authors to become defensive against the reviewers' advice. This leads them to argue against each point in their memo rather than to learn constructively from the reviews about how to improve the research. Another mistake is for authors to ignore advice that the editors explicitly provide. The editors are making the final decision on your manuscript so you cannot afford to alienate them. You should be aware of the journal's approach to handling manuscripts with R&R decisions. Some journals send the manuscript to the original reviewers plus a new reviewer, while other journals either send it back only to the original reviewers or make an inhouse editorial decision. These procedures can dramatically influence your chances for success at the R&R stage. If the paper is sent to a new reviewer, you should expect another R&R decision to be very likely.[5]

Getting a revise and resubmit decision is exciting for an author but also a daunting process when one sees how many revisions might be expected. You have to determine how to strike a balance between defending your ideas and revising your work in response to the criticisms you have received in the peer review process. My observation is that authors who are open to criticism and can learn from reviewers' suggestions are more successful in converting R&Rs to publications.

[4]Djupe's (2015, 346-7) survey of APSA members shows that 90% of tenured or tenure-track faculty reviewed for a journal in the past calendar year, with the average number of reviews varying by rank (assistant professors-5.5, associate professors-7, and full professors-8.3). In an analysis of review requests for the *American Political Science Review*, Breuning et al. (2015) find that while 63.6% of review requests are accepted, scholars declining the journal's review requests often note that they are too busy with other reviews. There is reasonable evidence that many political scientists feel overburdened by reviews, although the extent to which extra appendices influence those attitudes is unclear from these studies.

[5]I have experienced this process myself at journals like *Journal of Peace Research* which send a paper to a new reviewer after the first R&R is resubmitted. I have only experienced three or more rounds of revisions on a journal article at journals that adopt this policy. My own personal preference as an editor is to make the decision in-house. I have a high standard for giving out R&Rs and thus feel qualified to make the final decision myself. One could argue, however, that by soliciting advice from new reviewers, the final published products might be better.

Peer Review Issues in our Discipline

Peer review is an essential part of our discipline for ensuring that political science publications are of the highest quality possible. In fact, I would argue that journal publishing, especially in the top journals in our field, is one of the few processes where a scholars' previous publication record or pedigree are not terribly important. My chances of getting accepted at *APSR* or *AJPS* have not changed over the course of my career. However, once I published a book with Cambridge University Press, I had many acquisitions editors asking me about ideas for future book publications. There are clearly many books in our discipline that have important influences on the way we think about political science research questions, but I would contend that journal publications are the ultimate currency for high caliber research given the high degree of difficulty for publishing in the best journals in our discipline.[6]

Having said that, I recognize that there are biases in the journal peer review process. One thing that surprised me in my career was how the baseline probability for publishing varied dramatically across different research areas. I worked in some areas where R&R or conditional acceptance was the norm and in other research areas where almost every piece was rejected.[7] For example, topics that have been very difficult for me to publish journal articles on include international law, international norms, human rights, and maritime conflicts. One of my early articles on the systemic democratic piece (Mitchell, Gates, and Hegre 1999) was published in a good IR journal despite all three reviewers being negative; the editor at the time (an advocate of the democratic peace himself) took at a chance on the paper. Papers I have written on maritime conflicts have been rejected at six or more journals before getting a single R&R decision. My work that crosses over into international law also tends to be rejected multiple times because satisfying both political science and international law reviewers can be difficult. Other topics I have written on have experienced more smooth sailing through journal review processes. Work on territorial and cross-border river conflicts has been more readily accepted, which is interesting given that maritime issues are also geopolitical in nature. Diversionary conflict and alliance scholars are quite supportive of each other's work in the review process. Other areas of my research agenda fall in between these extremes. My empirical work on militarized conflict (e.g. the issue approach) or peaceful conflict management (e.g. mediation) can draw either supportive or really tough reviewers, a function I believe of the large number of potential reviewers in these fields. I have seen similar patterns in advising PhD students. Some students who were working in emerging topics like civil wars or terrorism found their work well-received as junior scholars, while others working on topics like foreign direct investment and foreign aid experienced more difficulties in converting their dissertation research into published journal articles.

Smaller and more insulated research communities can be helpful for junior scholars if the junior members are accepted into the group, as the chances for publication can be higher. On the other hand, some research areas have a much lower baseline publication rate. Anything that is interdisciplinary in my experience lowers the probability of success, which is troubling from a diversity perspective given the tendency for women and minority scholars to be drawn to interdisciplinary research. As noted above, I have also observed that certain types of work (e.g. empirical conflict work or research on gender) face more obstacles in the review process because there are a larger number of potential reviewers, which also increases the risks that at least one person will dislike your research. In more insulated communities, the number of potential reviewers is small and they are more likely to agree on what constitutes good research. Junior scholars may not know the baseline probability of success in their research area, thus it is important to talk with senior scholars about their experiences publishing on specific topics. I typically recommend a portfolio strategy with journal publishing, where junior scholars seek to diversify their substantive portfolio, especially if the research community for their dissertation project is not receptive to publishing their research.

I also think that journal editors have a collective responsibility to collect data across research areas and determine if publication rates vary dramatically. We often report on general subfield areas in annual journal reports, but we do not typically break down the data into more fine-grained research communities. The move to having scholars click on specific research areas for reviewing may facilitate the collection of this information. If reviewers' recommendations for R&R or acceptance vary across research topics, then having this information would assist new journal editors in making editorial decisions. Once we collect this kind of data, we could also see how these intra-community reviewing patterns influence the long term impact of research fields. Are broader communities with lower probabilities of publication success more effective in the long run in terms of garnering citations to the research? We need additional data collection to assess my hypothesis that baseline publication rates vary across substantive areas of our discipline.

We also need to remain vigilant in ensuring representation of women and minority scholars in political science

[6]Clearly there are differences in what can be accomplished in a 25-40 page journal article versus a longer book manuscript. Books provide space for additional analyses, in-depth case studies, and more intensive literature reviews. However, many books in my field that have been influential in the discipline have been preceded by a journal article summarizing the primary argument in a top ranked journal.

[7]This observation is based on my own personal experience submitting articles to journals and thus qualifies as more of a hypothesis to be tested rather than a settled fact.

journals. While women constitute about 30% of faculty in our discipline (Mitchell and Hesli 2013), the publication rate by women in top political science journals is closer to 20% of all published authors (Bruening and Sanders 2007). Much of this dynamic is driven by a selection effect process whereby women spend less time on research relative to their male peers and submit fewer papers to top journals (Allen 1998; Link, Swann, and Bozeman 2008; Hesli and Lee 2011). Journal editors need to be more proactive in soliciting submissions by female and minority scholars in our field. Editors may also need to be more independent from reviewers' recommendations, especially in low success areas that comprise a large percentage of minority scholars. It is disturbing to me that the most difficult areas for me to publish in my career have been those that have the highest representation of women (even though it is still small!). We cannot know whether my experience generalizes more broadly without collection of data on topics for conference presentations, submissions of those projects to journals, and the average "toughness" of reviewers in such fields. I believe in the peer review process and I will continue to provide public goods to protect it. I also believe that we need to determine if the process is generating biases that influence the chances for certain types of scholars or certain types of research to dominate our best journals.

References

Allen, Henry L. 1998. "Faculty Workload and Productivity: Gender Comparisons." In *The NEA Almanac of Higher Education.* Washington, DC: National Education Association.

Breuning, Marijke, Jeremy Backstrom, Jeremy Brannon, Benjamin Isaak Gross, and Michael Widmeier. 2015. "Reviewer Fatigue? Why Scholars Decline to Review their Peers' Work." *PS: Political Science & Politics* 48 (4): 595-600.

Breuning, Marijke, and Kathryn Sanders. 2007. "Gender and Journal Authorship in Eight Prestigious Political Science Journals." *PS: Political Science & Politics* 40 (2): 347-51.

Djupe, Paul A. 2015. "Peer Reviewing in Political Science: New Survey Results." *PS: Political Science & Politics* 48 (2): 346-52.

Hesli, Vicki L., and Jae Mook Lee. 2011. "Faculty Research Productivity: Why Do Some of Our Colleagues Publish More than Others?" *PS: Political Science & Politics* 44 (2): 393-408.

Link, Albert N., Christopher A. Swann, and Barry Bozeman. 2008. "A Time Allocation Study of University Faculty." *Economics of Education Review* 27 (4): 363-74.

Miller, Beth, Jon Pevehouse, Ron Rogowski, Dustin Tingley, and Rick Wilson. 2013. "How To Be a Peer Reviewer: A Guide for Recent and Soon-to-be PhDs." *PS: Political Science & Politics* 46 (1): 120-3.

Mitchell, Sara McLaughlin, Scott Gates, and Håvard Hegre. 1999. "Evolution in Democracy-War Dynamics." *Journal of Conflict Resolution* 43 (6): 771-92.

Mitchell, Sara McLaughlin and Vicki L. Hesli. 2013. "Women Don't Ask? Women Don't Say No? Bargaining and Service in the Political Science Profession." *PS: Political Science & Politics* 46 (2): 355-69.

Offering (Constructive) Criticism When Reviewing (Experimental) Research

Yanna Krupnikov
Stony Brook University
yanna.krupnikov@stonybrook.edu

Adam Seth Levine
Cornell University
asl22@cornell.edu

No research manuscript is perfect, and indeed peer reviews can often read like a laundry list of flaws. Some of the flaws are minor and can be easily eliminated by an additional analysis or a descriptive sentence. Other flaws often stand – at least in the mind of the reviewer – as a fatal blow to the manuscript.

Identifying a manuscript's flaws is part of a reviewer's job. And reviewers can potentially critique every kind of research design. For instance, they can make sweeping claims that survey responses are contaminated by social desirability motivations, formal models rest on empirically-untested assumptions, "big data" analyses are not theoretically-grounded, observational analyses suffer from omitted variable bias, and so on.

Yet, while the *potential* for flaws is ever-present, the key for reviewers is to go beyond this potential and instead ascertain whether such flaws *actually* limit the contribution of the manuscript at hand. And, at the same time, authors need to communicate why we can learn something useful and interesting from their manuscript despite the research design's potential for flaws.

In this essay we focus on one potential flaw that is often mentioned in reviews of behavioral research, especially research that uses experiments: critiques about external validity based on characteristics of the sample.

In many ways it is self-evident to suggest that the sample (and, by extension, the population that one is sampling from) is a pivotal aspect of behavioral research. Thus it is not surprising that reviewers often raise questions not only about the theory, research design, and method of data analysis, but also the sample itself. Yet critiques of the sample are often stated in terms of potential flaws – that is, they are based on the *possibility* that a certain sample could affect the conclusions drawn from an experiment rather than stating how the author's *particular* sample affects the inferences that we can draw from his or her *particular* study.

Here we identify a concern with certain types of sample-focused critiques and offer recommendations for a more constructive path forward. Our goals are complimentary and twofold: first, to clarify authors' responsibilities when justifying the use of a particular sample in their work and, second, to offer constructive suggestions for how reviewers should evaluate these samples. Again, while our arguments could apply to all manuscripts containing behavioral research, we pay particular attention to work that uses experiments.

What's the Concern?

Researchers rely on convenience samples for experimental research because it is often the most feasible way to recruit participants (both logistically and financially). Yet, when faced with convenience samples in manuscripts, reviewers may bristle. At the heart of such critiques is often the concern that the sample is too "narrow" (Sears 1986). To argue that a sample is narrow means that the recruited participants are homogenous in a way that differs from other populations to which authors might wish to generalize their results (and in a way that affects how participants respond to the treatments in the study). Although undergraduate students were arguably the first sample to be classified as a "narrow database" (Sears 1986), more recently this label has been applied to other samples, such as university employees, residents of a single town, travelers at a particular airport, and so on.

Concerns regarding the narrowness of a sample typically stem from questions of external validity (Druckman and Kam 2011). External validity refers to whether a "causal relationship holds over variations in persons, settings, treatments and outcomes" (Shadish, Cook and Campbell 2002, 83). If, for example, a scholar observes a result in one study, it is reasonable to wonder whether the same result could be observed in a study that altered the participants or slightly adjusted the experimental context. While the sample is just one of many aspects that reviewers might use when judging the generalizability of an experiment's results – others might include variations in the setting of the experiment, its timing, and/or the way in which theoretical entities are operationalized – sample considerations have often proved focal.

At times during the review process, the type of sample has become a "heuristic" for evaluating the external validity of a given experiment. Relatively "easy" critiques of the sample – those that dismiss the research simply because they involve a particular convenience sample – have evolved over time. A decade ago such critiques were used to dismiss experiments altogether, as McDermott (2002, 334) notes: "External validity...tend[s] to preoccupy critics of experiments. This near obsession… tend[s] to be used to dismiss experiments." More recently, Druckman and Kam (2011) noted such concerns were especially likely to be directed toward experiments with student samples: "For political scientists who put particular emphasis on generalizability, the use of student participants often constitutes a critical, and according to some reviewers, fatal problem for experimental studies." Even more recently, reviewers lodge this critique against other convenience samples such as those from Amazon's Mechanical Turk.

Note that, although they are writing almost a decade apart, both McDermott and Druckman and Kam are observing the same underlying phenomenon: reviewers dismissing experimental research simply because it involves a particular sample. The review might argue that the participants (for example, undergraduate students, Mechanical Turk workers, or any other convenience sample) are *generally* problematic, rather than arguing that they pose a problem for the *specific* study in the manuscript.

Such general critiques that identify a broad potential problem with using a certain sample can, in some ways, be more even damning than other types of concerns that reviewers might raise. An author could address questions of analytic methods by offering robustness checks. In a well-designed experiment, the author could reason through questions of alternative explanations using manipulation checks and alternative measures. When a review suggests that the core of the problem is that a sample is generally "bad", however, the reviewer is (indirectly) stating that readers cannot glean much about the research question from the authors' study and that the reviewer him/herself is unlikely to be convinced by any additional arguments the author could make (save a new experiment on a different sample).

None of the above is to suggest that critiques of samples should *not* be made during the review process. Rather, we believe that they should adhere to a similar structure as concerns that reviewers might raise about other parts of a manuscript. Just as reviewers evaluate experimental treatments and measures within the context of the authors' hypotheses and specific experimental design, evaluations of the sample also benefit from being experiment-specific. Rather than asking "is this a 'good' or 'bad' sample?", we suggest that reviewers ask a more specific question: "is this a 'good' or 'bad' sample given the author's research goals, hypotheses, measures, and experimental treatments?"

A Constructive Way Forward

When reviewing a manuscript that relies on a convenience sample, reviewers sometimes dismiss the results based on the potential narrowness of a sample. Such a dismissal, we argue, is a narrow critique. The narrowness of a sample certainly *can* threaten the generalizability of the results, but it does not do so unconditionally. Indeed, as Druckman and Kam (2011) note, the narrowness of a sample is limiting if the sample lacks variance on characteristics that affect the way a participant responds to the particular treatments in a given study.

Consider, for example, a study that examines the attitudinal impact of alternative ways of framing health care policies. Suppose the sample is drawn from the undergraduate population at a local university, but the researcher argues (either implicitly or explicitly) that the results can help us understand how the broader electorate might respond to these alternative framings.

In this case, one potential source of narrowness might stem from personal experience. We might (reasonably) assume that undergraduate students are highly likely to have experience interacting with a doctor or a nurse (just like non-undergraduate adults). Yet, they are perhaps far less likely to have experience interacting with health insurance administrators (unlike non-undergraduate adults). When might this difference threaten the generalizability of the claims that the author wishes to make?

The answer depends upon the specifics of the study. If we believe that personal experience with health care providers and insurance administrators does not affect how people respond to the treatments, then we would not have reason to believe that the narrowness of the undergraduate sample would threaten the authors' ability to generalize the results. If instead we only believe that experience with a doctor or nurse may affect how people respond to the treatments (e.g. perhaps how they comprehend the treatments, the kinds of considerations that come to mind, and so on) then again we would not have reason to believe that the narrowness of the undergraduate sample would threaten the ability to generalize. Lastly, however, if we also believe that experience with insurance administrators affects how people respond to the treatments, then that would be a situation in which the narrowness might limit the generalizability of the results.

What does this mean for reviewers? The general point is that, even if we have reason to believe that the results would differ if a sample were drawn from a different population, this fact does not render the study or its results entirely invalid. Instead, it changes the conclusions we can draw. Returning to the example above, a study in which experience with health insurance administrators affects responses still offers some political implications about health policy messages. But (for example) its scope may be limited to those with very little experience interacting with insurance administrators.

It's worth noting that in some cases narrowness might be based on more abstract, psychological factors that apply across several experimental contexts. For instance, perhaps reviewers are concerned that undergraduates are narrow because they are both homogeneous and different in their reasoning capacity from several other populations to which authors often wish to generalize. In that case, the most constructive review would explain *why* these reasoning capacities would affect the manuscript's conclusions and contribution.

More broadly, reviewers may also consider the researcher's particular goals. Given that some relationships are otherwise difficult to capture, experimental approaches often offer the best means for identifying a "proof of concept" – that is, whether under theorized conditions a "particular behavior emerges" (McKenzie 2011). These types of "proof of concept" studies may initially be performed only in the laboratory and often with limited samples. Then, once scholars observe some evidence that a relationship exists, more generalizable studies may be carried out. Under these conditions, a reviewer may want to weigh the possibility of publishing a "flawed" study against the possibility of publishing no evidence of a particularly elusive concept.

What does this mean for authors? The main point is that it is the author's responsibility to clarify why the sample is appropriate for the research question and the degree to which the results may generalize or perhaps be more limited. It is also the author's responsibility to explicitly note why the result is important even despite the limitations of the sample.

What About Amazon's Mechanical Turk?

Thus far we have (mostly) avoided mentioning Amazon's Mechanical Turk (MTurk). We have done so deliberately, as MTurk is an unusual case. On the one hand, MTurk provides a platform for a wide variety of people to participate in tasks such as experimental studies for money. One result is that MTurk typically provides samples that are much more heterogeneous than other convenience samples and are thus less likely to be "narrow" on important theoretical factors (Huff and Tingley 2015). These participants often behave much like people recruited in more traditional ways (Berinsky, Huber and Lenz 2012). On the other hand, MTurk participants are individuals who were somehow motivated to join the platform in the first place and over time (due to the potentially unlimited number of studies they can take) have become professional survey takers (Krupnikov and Levine 2014; Paolacci and Chandler 2014). This latter characteristic in particular suggests that MTurk can produce an unusual set of challenges for both authors and reviewers during the manuscript review process.

Much as we argued that a narrow sample is not in and of itself a reason to advocate for a manuscript's rejection (though the interaction between the narrowness of the sample and the author's goals, treatments and conclusions may provide such a reason), so too when it comes to MTurk we believe that this recruitment approach does not provide *prima facie* evidence to advocate rejection.

When using MTurk samples, it is the author's responsibility to acknowledge and address any potential narrowness of the sample that might stem from the sample. It is also the author's responsibility to design a study that accounts for the fact that MTurkers are professionalized participants (Krupnikov and Levine 2014) and to explain why a particular study is not limited by the characteristics that make MTurk unusual. At the same time, we believe that reviewers should avoid using MTurk as an unconditional heuristic for rejection and instead should always consider the relationship between treatment and sample in the study at hand.

Conclusions

We are not the first to note that reviewers can voice concerns about experiments and/or the samples used in experiments. These types of sample critiques may often seem unconditional, as in: there is no amount of information that the author could offer that could lead the reviewer to reconsider his or her position on the sample. Put another way, the sample type is used as a heuristic, with little consideration of the specific experimental context in the manuscript.

We are not arguing that reviewers should never critique samples. Rather, our argument is that the fact that researchers chose to recruit a convenience sample from the population of undergraduates at a local university, the population of MTurk workers, and so on is not a justifiable reason *on its own* for recommending rejection of a paper. Rather, the validity of the sample depends upon the author's goals, the experimental design, and the interpretation of the results. The use of undergraduate students may have few limitations for one experiment but may prove largely crippling for another one. And, echoing Druckman and Kam (2011), even a nationally-representative sample is no guarantee of external validity.

The reviewer's task, then, is to examine how the sample interacts with all the other components of the manuscript. The author's responsibility, in turn, is to clarify such matters. And, in both cases, both the reviewer and the author should acknowledge that the only way to truly answer questions about generalizability is to continue examining the question in different settings as part of an ongoing research agenda (McKenzie 2011).

Lastly, while we have focused on a common critique of experimental research, this is just one example of a broader phenomenon. All research designs are imperfect in one way or another, and thus the potential for flaws is always present. Constructive reviews should evaluate such flaws in the context of the manuscript at hand, and then decide if the manuscript credibly contributes to our knowledge base. And, similarly, authors are responsible for communicating the value of their manuscript despite any potential flaws stemming from their research design.

References

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351-68.

Druckman, James N., and Cindy D. Kam. 2011. "Students as Experimental Participants: A Defense of the "Narrow Data Base."" In *Cambridge Handbook of Experimental Political Science*, eds. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. Cambridge: Cambridge University Press, 41-57.

Huff, Connor, and Dustin Tingley. 2015. ""Who are these people?" Evaluating the Demographic Characteristics and Political Preferences of MTurk Survey Respondents." *Research and Politics* 2 (3) DOI: 10.1177/2053168015604648.

Krupnikov, Yanna, and Adam Seth Levine. 2014. "Cross-Sample Comparisons and External Validity." *Journal of Experimental Political Science* 1 (1): 59-80.

McDermott, Rose. 2002. "Experimental Methodology in Political Science." *Political Analysis* 10 (4): 325-42.

McKenzie, David. 2011. "A Rant on the External Validity Double Double-Standard." May 2. *Development Impact: The World Bank*. `http://blogs.worldbank.org/impactevaluations/a-rant-on-the-external-validity-double-double-standard` (February 1, 2016).

Paolacci, Gabriele, and Jesse Chandler. 2014. "Inside the Turk: Understanding Mechanical Turk as a Participant Pool." *Current Directions in Psychological Science* 23 (3): 184-8.

Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51: 515-30.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. [2001] 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company.

2015 TPM Annual Most Viewed Post Award Winner

Justin Esarey
Rice University
justin@justinesarey.com

On behalf of the editorial team at *The Political Methodologist*, I am proud to announce the 2015 winner of the Annual Most Viewed Post Award: Nicholas Eubank of Stanford University! Nick won with his very informative post "A Decade of Replications: Lessons from the Quarterly Journal of Political Science (Eubank 2014)." This award entitles the recipient to a line on his or her curriculum vitae and one (1) high-five from the incumbent *TPM* editor (to be collected at the next annual meeting of the Society for Political Methodology).[1]

The award is determined by examining the total accumulated page views for any piece published between December 1, 2014 and December 31, 2015; pieces published in December are examined twice to give them a fair chance to garner page views. The page views for December 2014 and calendar year 2015 are shown below; orange hash marks next to the post indicate that it was published during the time period (and thus eligible to receive the award).

Stats for December 2014

Title		Views
Making High-Resolution Graphics for Academic Publishing	View	1,709
A Decade of Replications: Lessons from the Quarterly Journal of Political Science	View	1,282
How tough should reviewers be?	View	626
Home page / Archives	View	545
Reproducibility and Transparency	View	100
The Use of Replication in Graduate Education and Training	View	99
Encountering your IRB: What political scientists need to know	View	89
Building and Maintaining R Packages with devtools and roxygen2	View	88
On the Replication of Experiments in Teaching and Training	View	73

Stats for 2015

Title		Views
Making High-Resolution Graphics for Academic Publishing	View	26,861
Home page / Archives	View	4,708
What courses do I need to prepare for a PhD in Political Science?	View	1,492
A Decade of Replications: Lessons from the Quarterly Journal of Political Science	View	1,476
A Checklist Manifesto for Peer Review	View	1,421
Building and Maintaining R Packages with devtools and roxygen2	View	1,206
Why Can't We Just Make Up Instrumental Variables?	View	996
Graphical Presentation of Regression Discontinuity Results	View	920
One Norm, Two Standards: Realizing Transparency in Qualitative Political Science	View	911
An Editor's Thoughts on the Peer Review Process	View	686
What is Peer Review For? Why Referees are not the Disciplinary Police	View	664
Implicit Bias and Why It Matters to the Field of Political Methodology	View	488
Why Do We Need Diversity in the Political Methodology Society?	View	471
Some code for estimating clustered SEs in mlogit models	View	409

Nick's contribution was viewed 2,758 times during the eligible time period, making him the clear winner for this year's award. Congratulations, Nick!

Our runner-up is Brendan Nyhan and his post "A Checklist Manifesto for Peer Review" which was viewed 1,421 times during the eligibility period. However, because Brendan's piece was published in December 2015, he'll be eligible for consideration for next year's award as well!

In related news, Thomas Leeper's post "Making High-Resolution Graphics for Academic Publishing" continues to dominate our viewership statistics; this post also won the 2014 *TPM* Most-Viewed Post Award. Although we don't have a formal award for him, I'm happy to recognize that Thomas's post has had a lasting impact among the readership and pleased to congratulate him for that achievement.

I am also happy to report that *The Political Methodologist* continues to expand its readership. Our 2015 statistics are here:

VIEWS	VISITORS
52,107	37,836

This represents a considerable improvement over last year's numbers, which gave us 43,924 views and 29,562 visitors. I'm especially happy to report the substantial increase

[1] A low-five may be substituted upon request.

in unique visitors to our site: over 8,000 new unique viewers in 2015 compared to 2014! Our success is entirely attributable to the excellent quality of the contributions published in *TPM*. So, thank you contributors!

References

Eubank, Nicolas. 2014. "A Decade of Replications: Lessons from the Quarterly Journal of Political Science." *The Political Methodologist* 22 (1): 16-9. http://thepoliticalmethodologist.com/2014/12/09/a-decade-of-replications-lessons-from-the-quarterly-journal-of-political-science/ (February 1, 2016).

Leeper, Thomas. 2013. "Making High-Resolution Graphics for Academic Publishing." *The Political Methodologist* 21 (1): 2-5. http://thepoliticalmethodologist.com/2013/11/25/making-high-resolution-graphics-for-academic-publishing/ (February 1, 2016).

Nyhan, Brendan. 2015. "A Checklist Manifesto for Peer Review." *The Political Methodologist* 23 (1): 4-6. http://thepoliticalmethodologist.com/2015/12/04/a-checklist-manifesto-for-peer-review/ (February 1, 2016).

Rice University
Department of Political Science
Herzstein Hall
6100 Main Street (P.O. Box 1892)
Houston, Texas 77251-1892

The Political Methodologist is the newsletter of the Political Methodology Section of the American Political Science Association. The support of the Department of Political Science at Rice University in helping to defray the editorial and production costs of the newsletter is gratefully acknowledged.

Subscriptions to *TPM* are free for members of the APSA's Methodology Section. Please contact APSA (202-483-2512) if you are interested in joining the section. Dues are $29.00 per year for students and $44.00 per year for other members. They include a free subscription to *Political Analysis*, the quarterly journal of the section.

Submissions to *TPM* are always welcome. Articles may be sent to any of the editors, by e-mail if possible. Alternatively, submissions can be made on diskette as plain ASCII files sent to Justin Esarey, 108 Herzstein Hall, 6100 Main Street (P.O. Box 1892), Houston, TX 77251-1892. LaTeX format files are especially encouraged.

TPM was produced using LaTeX.

THE SOCIETY FOR
POLITICAL METHODOLOGY
ad indicia spectate