Comparative Political Studies
2016, Vol. 49(13) 1667–1703
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0010414016655539
cps.sagepub.com
SAGE

Can Results-Free Review Reduce Publication Bias? The Results and Implications of a Pilot Study

Michael G. Findley[1], Nathan M. Jensen[1], Edmund J. Malesky[2], and Thomas B. Pepinsky[3]

Abstract

In 2015, *Comparative Political Studies* embarked on a landmark pilot study in research transparency in the social sciences. The editors issued an open call for submissions of manuscripts that contained no mention of their actual results, incentivizing reviewers to evaluate manuscripts based on their theoretical contributions, research designs, and analysis plans. The three papers in this special issue are the result of this process that began with 19 submissions. In this article, we describe the rationale for this pilot, expressly articulating the practices of preregistration and results-free review. We document the process of carrying out the special issue with a discussion of the three accepted papers, and critically evaluate the role of both preregistration and results-free review. Our main conclusions are that results-free review encourages much greater attention to theory and research design, but that it raises thorny problems about how to anticipate and interpret null findings. We also observe that as currently practiced, results-free review has a particular affinity with experimental and cross-case

[1]University of Texas at Austin, TX, USA
[2]Duke University, Durham, NC, USA
[3]Cornell University, Ithaca, NY, USA

Corresponding Author:
Michael G. Findley, Department of Government, University of Texas at Austin, 3.108 Batts, Department of Gov, Austin, TX 78712, USA.
Email: mikefindley@utexas.edu

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

methodologies. Our lack of submissions from scholars using qualitative or interpretivist research suggests limitations to the widespread use of results-free review.

Keywords
experimental research, quantitative methods, qualitative methods, results-free review, transparency, preregistration

Introduction

In the past decade, political science has witnessed a growing movement for greater transparency in research. Prominent examples include efforts by the Evidence in Governance and Politics Network (EGAP; www.egap.org), the Berkeley Initiative for Transparency in the Social Sciences (BITSS; www.bitss.org), a recent symposium on transparency in qualitative methods (Moravcsik, 2014), and the recent Data Access and Research Transparency (DART) statement signed by the editors of 27 leading journals (http://www.dartstatement.org/).

Although there are varied objectives driving the shift toward greater transparency, one of the key motivations is to avoid *publication bias*, which can emerge as a result of a peer-review process that privileges the significance of results over their theoretical contribution, research design, quality of the data and analysis, and even the importance of the motivating research question. So long as the significance of results is the overriding concern among editors and reviewers, authors will have few incentives to report all of the empirical tests they conduct. Publication bias can manifest itself through bias in individual studies, but aggregated across studies an overall bias manifests itself in the scholarly record in a given area. Moreover, it can lead to serious questions about the overall quality of research in the field, as evidenced by the recent crisis in psychology (Open Science Collaboration, 2015; Yong, 2012).

A potentially simple and yet powerful way to mitigate publication bias is for journals to commit to publish manuscripts without any knowledge of the actual findings. Authors might submit sophisticated research designs that serve as a registration of what they intend to do.[1] Or they might submit already completed studies for which any mention of results is expunged from the submitted manuscript. Reviewers would carefully analyze the theory and research design of the article. If they found that the theoretical contribution was justifiably large and the design an appropriate test of the theoretical logic, then reviewers could recommend publication regardless of the final outcome of the research. In theory, this could mitigate publication bias (see Nyhan, 2014).

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Implementing such a system is challenging, and full of uncertainty. This Special Issue of *Comparative Political Studies* (*CPS*) helps to assess the potential benefits and costs associated with new models of the publication process by studying how this particular model works in practice. In so doing, we shed new light on the transparency debate in the social sciences. We consider it to be self-evidently true that transparency should be one central objective in contemporary social science, but what are the costs and benefits of different transparency approaches and in what ways would current publication practices have to change to accommodate a results-free review model? Some critics of results-free review, for example, may worry that it will inevitably lead to journals full of null results, and to projects that are less theoretically innovative and path breaking than would otherwise be possible. In other words, journals would receive and publish "boring" work. Does results-free review commit scholars to carry out projects that are unfeasible, or dissuade creative dialogue between theory and data? How will manuscript referees respond to manuscripts without results or conclusions? These questions cannot be settled in the abstract.

We investigated these questions through a special issue on research transparency. The goal of this special issue was to consider papers that fit within the mandate of *CPS*, but were submitted as standalone designs or completed papers without any of the results reported. Thus, our special issue is not on a substantive theme or topic, but is rather defined by the process whereby authors submitted manuscripts and referees reviewed them. As special issue editors, we were involved in the entire process, which meant that we observed the submissions through to acceptance. Like the peer reviewers, though, we too never once saw the results of any of the papers until the final stage, well after publication decisions were finalized.

We created the call for papers (CFP) and advertised broadly, we observed what kinds of submissions we received, added our own evaluations of the manuscripts, decided which pieces to desk reject or send out for review, selected reviewers, received reviewer comments, and made final recommendations to the *CPS* standing editors. Of course, we worked closely with the standing editors throughout the entire process. Our close involvement gave us helpful insights into how results-free review works in practice, which we share here.

Based on this experience, in this introductory essay, we make three key observations about results-free peer review in practice. First, contrary to fears that greater emphasis on transparency creates more incentives for clever research designs and methodological perfection, reviewers placed an overwhelming emphasis on theoretical consistency and substantive importance. In this regard, results-free review worked better than we could have

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

hoped in incentivizing theory and research design over narrow concerns about novelty of methodology or empirical causal identification. Of course, our pilot was not a direct comparison between results-free review and the same exact papers undergoing standard reviews, so we lack the counterfactual to make definitive conclusions about the benefits of the process. We can say with confidence, however, that reviewers in our pilot were explicitly concerned about well-articulated theories and null results that moved literatures forward scientifically, and were not tolerant of long lists of ambiguous hypotheses to be tested (what we refer to as hypothesis trolling). Atheoretical and "boring" work stood very little chance of publication in this pilot. Relatedly, we hasten to add that the overall quality of the reviews for the special issue were quite strong. Indeed, it appears that by needing to engage the theory and hypotheses, reviewers could not simply nitpick over the credibility of statistical results. Improved engagement by reviewers is one argument that editors may consider when allowing results-free review as a submission option.

Second, it was nevertheless immensely challenging for reviewers and authors alike to argue coherently about the proper role of null findings, often referred to tellingly as "non-results." The challenges for current practices in this regard are steeper than we had anticipated, and speak to general debates about null-significance hypothesis testing and the relationship between theory, data, and models (Clarke & Primo, 2012).

And, third, results-free review has a particular affinity for certain methodologies, reflected in the types of submissions we received for this special issue. In particular, our submissions were almost exclusively experiments and observational studies that were testing general propositions using cross-case inferential techniques. Each of these three observations, we argue, has substantial implications for social science in general, for comparative politics in particular, and also for contemporary debates about transparency.

The rest of this essay proceeds as follows. In the next section, we provide an overview of publication bias and then consider the premise that results-free peer review could be a potential solution to the problem. We then briefly outline the procedures that we followed in our special issue. The subsequent section discusses what we learned—the importance of theory, null findings, and methodological affinities. Next, we summarize the findings in the three articles that successfully completed the peer-review process and which appear in the special issue, noting ways in which the process behind each reflects these general concerns. A final section provides some practical considerations about how to manage a results-free process in a top-flight journal.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

The Problem of Publication Bias

Before diving into the details of our special issue, we take a step back and review the general challenge of publication bias. How do we know such bias exists in political science journals? What factors have driven it? What steps have been taken thus far to address it? And how successful have those steps been? In this section, we answer these questions before discussing how results-free review might assist in the battle against publication bias.

What Is Publication Bias?

In its most basic form, publication bias exists when a set of published studies is not representative of all available or possible studies. There are myriad reasons for a non-representative set of available studies. In much scientific work, publication bias is most pronounced when publication decisions are based on the realized outcomes of a study—typically statistical significance of a result—rather than the merits of the approach and design (Dickersin, 1990; Humphreys, de la Sierra, & van der Windt, 2013; Sterling, 1959).

The problem of publication bias is not complicated, but it is rampant and consequential. One goal in political science (and social science more generally) is the correct measurement of causal effects. If a study is carried out correctly, then the results should matter regardless of whether they confirm preexisting hypotheses about those causal effects. Indeed, null results from a well-designed study are just as meaningful as strong positive or negative effects. To take one prominent example from comparative politics, what if there is no causal relationship between political culture and democratic rule? If this is so, then this guides our understanding as a discipline of the origins of democratic regimes, and may also serve to guide policy makers who wish to promote democracy. But what if results that show no link between values and political regimes are less likely to be published than are results that support the existence of such a relationship? If so, then even for a question in which there is vigorous debate between findings and null findings (see, recently, Welzel & Inglehart, 2009), the *published* evidence will imply that the claim has more empirical support than it does.

Unfortunately, existing publication practices in social science and other disciplines privilege strong (i.e., statistically significant and substantively large) positive or negative findings, thus making null effects less likely to emerge. As such, scholars have compelling incentives to engage in data fishing (Humphreys et al., 2013) to obtain results that will be publishable. What is especially problematic is that even if individual scholars find ways to precommit to not engaging in data fishing, the publication process could lead to

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

bias. Studies with null results may not survive the publication process due to reviewer and editorial decisions, as other studies with large (and perhaps more counter-intuitive results) are more likely to be published. And if those published results are not representative of the actual distribution of causal effects (or lack thereof) in the real world, then publication bias exists and skews our knowledge base as well as any public policy that results from a given corpus of studies. Even in a world of angels writing research papers, the devil may still hide in the peer-review process.

How prevalent is publication bias? In a recent examination of the *American Political Science Review* and the *American Journal of Political Science*, Gerber and Malhotra (2008) conducted an extensive survey and test the hypothesis of whether publication bias exists. In their own staggering words, "we can reject the hypothesis of no publication bias at the 1 in 32 billion level" (p. 313). In a more recent study, Franco, Malhotra, and Simonovits (2014) document publication bias across the known population of studies utilizing the Time-Sharing Experiments in the Social Sciences (TESS) program and demonstrate that fielded survey experiments with null findings are substantially less likely to be published than those with significant effects, and the principal investigators themselves admit to abandoning such "unsuccessful" projects. Sadly, these results confirm what others have found across the social sciences (see, for example, Dickersin, 1990; Gerber, Malhotra, Dowling, & Doherty, 2010; Glewwe & Kremer, 2006; Ioannidis, 1998).

The implications of publication bias for the social sciences may be more consequential than distorting the findings in academic journals. The consequences of incorrect findings, championed as scientific evidence, are obvious in medicine, and can lead to incorrect diagnosis or treatment, such as decades of demonizing saturated fats despite clear, unpublished experimental evidence to the contrary (O'Connor, 2016). Similar problems may confront the social sciences.[2] Returning to the democracy example from above, academic research by Finkel, Pérez-Liñán, and Seligson (2007) has informed democracy promotion decisions by the U.S. Agency for International Development (USAID), and yet the study's few challengers find little to no evidence that the statistically significant results in Finkel et al. (2007) hold. Although far from a settled debate, these studies illustrate that research has the potential to influence the decisions of well-meaning practitioners and policy makers in government, even before scientific consensus is reached. And, needless to say, public policy decisions in the social sciences can have large and lasting impacts on peoples' lives.

One prominent example from political science serves to show the most extreme case. Most political scientists are familiar with the controversy surrounding a paper published by LaCour and Green (2014) on the impact of

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

canvasing on support for gay marriage. This article is notable for two reasons. First, it was retracted due to alleged fraud in essentially every aspect of the project including false statements about grant funding, compensation of subjects, preregistration of hypotheses, Human Subjects approval, and even the very data collection itself.[3] Second, what may have made this article especially interesting in the first place was the massive size and persistence of the impact of gay canvasing on support for gay marriage that stood in sharp contrast to nearly all existing studies on persuasion.[4]

Our special issue has little to say about outright fraud in research. But the counterfactual we would like *CPS* readers to consider is as follows: What would have happened to this paper if it had been reviewed without results? Would this have led to additional scrutiny of the paper that would have uncovered the fraud? Perhaps. More important for our exercise, the merits of publishing this article would not have been based on its splashy results. Scrutiny of how this research design relates to existing work in the field, data collection efforts, and an analysis plan would have been central to the success or failure of the paper.

This is clearly an extreme case of fraud, and our counterfactual is speculative. But it is helpful to recognize that there are two distinct drivers of publication bias. The first source is the career-oriented motivations of individual authors. Job placement and tenure decisions can generate incentives for (a) prioritizing work in the pipeline that has "significance stars" next to key coefficients, which was the key lesson of Franco et al. (2014); (b) choking unresponsive data with a barrage of specification choices and subsample analyses until the data confess (Nuzzo, 2014);[5] and, rarely but it happens, (c) outright manipulation or fraud.[6]

The second source of bias occurs when reviewers and editors evaluate the publication merits of null results. The burden of proof appears to be higher for null rather than significant results, because reviewers are forced to decide whether incorrect theory or a problematic research design generated the insignificant result. Recently, one of the editors of this special issue received a revise and resubmit decision from a prominent journal with encouragement to abandon null results. The reviewers cited theoretical deficiencies, leaving a difficult decision of whether to push back and keep the null results or drop them at the behest of the reviewers. Academic incentives for junior faculty or grad students likely result in following the reviewers in cases like this.

First and foremost, null results often call into question the larger theoretical enterprise of the paper. Skeptical reviewers might give the benefit of the doubt to an implausible theory that, despite reviewer misgivings, yielded observable implications that were tested and identified with significant findings. The same generosity would not be provided to null findings. If a

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

reviewer never believed that sunspots influenced social movements to begin with, why would an empirical test that finds no evidence of such a relationship be worth publishing? Gelman and Carlin (2014) describe several examples of theoretically implausible but highly provocative findings that were indeed published because of their statistical significance, and show that reasonable calculations of the likely effect sizes in the studies in question imply that the reported effects are massive overestimates—and very possibly have the wrong sign.

The second issue is empirical, and the frequentist language for hypothesis testing is helpful for elucidating the problem. With a null finding, we "fail to reject" a null hypothesis, we do not "disprove" the alternative. Why does this matter? As a thought experiment, imagine an accounting ledger of research design flaws that might bias in favor of rejecting the null hypothesis when it is correct (Type I error) or failing to reject the null hypothesis when the alternative is, in fact, correct (Type II error). A number of mistakes can lead to biased coefficients and Type I errors, including simultaneity, biased selection into treatment, systematic measurement error in the independent variable, omitted variable bias, and unobserved heterogeneity. The list for Type II errors, however, includes all of those issues and a few more that either bias coefficients to zero or increase inefficiency, including insufficient power, stochastic measurement error in the independent variable (leading to white noise), and stochastic measurement error in the dependent variable (leading to attenuation bias). Thus, on a simple accounting basis, papers with null findings have to overcome a greater set of inferential obstacles than those with significant coefficients.

The problem actually goes deeper, however, as even when a flaw is apparent, the burden of demonstrating the robustness of findings to a potential correction is easier for Type I errors. First of all, the most common stratagem for side-stepping a research design flaw that would bias against a significant coefficient is not available to scholars with null findings. Authors aware of measurement error that creates noise and increases their standard errors, for example, will often claim that they obtained significance despite the stochastic measurement error. Similarly, authors aware of omitted variable bias can argue that the excluded variable would have likely biased the coefficient on their key causal variable toward zero. The fact that they still found a significant effect despite the bias indicates that their main effects would even be stronger if they had a properly specified regression.

The appeal to the persistence of stars in the presence of bias is not available for null findings. Similarly, scholars with significant findings can demonstrate the insensitivity of their significant coefficient to multiple measures, introduction of confounders, and specification choices, "Despite multiple

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

attempts, I could not make those stars disappear." This approach, however, rings hollow when the stars were never there. Adding more models with null findings only appears to reinforce that the alternative specifications have not addressed the underlying design flaw.

All that said, there are techniques to mitigate the threat of null findings, distinguishing between design problems and deeper concerns. The most common of these is assessment of a study's power. By ensuring adequate statistical power, a study attempts to avoid Type II errors. As commonly defined, statistical power is the probability of correctly rejecting the null hypothesis, or, alternatively, 1 minus the probability of committing a Type II error. Put this way, if a study is sufficiently powered, then authors can argue that they have solved at least some design issues that could have yielded the null result and therefore offer greater confidence that the null result is meaningful.[7]

The specific point on statistical power is indicative of a more general point about the quality of the research question, theory, and design. It may be that few null results get published precisely because the quality of the research that produced those results was so low. Given the set of possible statistical relationships that could be explored, scholars typically begin by theorizing about those that ought to be related, thereby leaving aside the investigation of true null relationships. Thus, a disproportionate share of significant findings could reflect on scholars carefully choosing questions and designing research appropriately, whereas null results could reflect on scholars' attempts to understand what should have been non-null relationships, but with low-quality research approaches.

Replication as a Solution to Publication Bias?

Any prescription for overcoming publication bias must first begin from the premise that null findings face a greater uphill battle for publication. This problem affects reviewers and is also clearly understood and appreciated by authors, which is why we suspect that researchers may consider efforts to publish null findings a fool's errand, and therefore do not even attempt to publish them, as Franco et al. (2014) showed.

Thus far, most efforts to address publication bias have focused on the first driver, the intentional actions of authors to produce work with p values below 0.05. To this end, efforts at producing greater transparency in research have thus far emphasized better replication practices. Most notably, the practice of making replication data available is increasingly common. The *Quarterly Journal of Political Science* (*QJPS*) has a staff member replicate all reported findings before publication. Other journals are beginning to follow suit, including the *American Journal of Political Science* and the *Journal of*

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Experimental Political Science. Other journals require posting replication files (e.g., *CPS, Journal of Politics, Journal of Conflict Resolution, Journal of Peace Research*, among others), and full platforms have made cataloging data both uniform and accessible (e.g., *Dataverse*: http://thedata.org/). These decentralized efforts on replication have culminated in a recent, but currently embattled, initiative by the American Political Science Association (APSA) to establish the DART standards, to which about 27 journals—including *CPS*—have agreed.

Perhaps due to the increased scrutiny, replication exercises have uncovered high-profile cases of academic fraud. As noted above, a study by Michael Lacour and Donald Green (2014; now retracted) was alleged to have used fabricated data through the process of replication (see Broockman, Kalla, & Aronow, 2014). One of the highest profile cases to hit political science, this scandal not only suggests the value of stronger transparency standards but also provides some initial validation that the system may at some level work in detecting unscrupulous behavior. Equally notable, but in psychology, is a fraud case by a Dutch researcher who falsified data and made up entire experiments (Carey, 2011). Although not academic fraud, not so long ago, a high-profile working paper by Reinhart and Rogoff (2010) found that government debt in excess of 80% of gross domestic product (GDP) had devastating consequences. Numerous politicians heralded this research as justification for fiscal policy reforms. And yet the authors provided little information on their coding rules and procedures for constructing their sample.[8] Eventually, it was uncovered that their results were driven by a questionable decision to drop some countries and a major Excel coding error (see Herndon, Ash, & Pollin, 2014). These are just a few examples, whereas organizations such as *Retraction Watch* document retractions (or discuss proposed retractions) of any kind and include a leader board of authors with the most retractions.[9]

Although data replication addresses some of the challenges of transparency, unfortunately it cannot set formidable standards against data fishing. For example, with few exceptions (see, for example, Nielsen, Findley, Candland, Davis, & Nielson, 2011), most replication studies provide only the data for the final set of results in a manuscript, thus leaving unknown the full set of data preparation operations conducted.

Preregistration as a Solution to Publication Bias

One possible way to increase transparency is through preregistration, which specifically requires that, prior to carrying out a study, scholars provide details about the research design of the study, how the study's data will be analyzed, as well as any potential conflict of interests with funders.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

These registries record all studies, including those that have been conducted and have not been submitted to journals. In this piece, we focus on the role of preregistration for submitted studies.

The field of medicine was the first to set up preregistration standards (De Angelis et al., 2004). The process of establishing mandatory preregistration in medicine was not easy; indeed, significant opposition contributed to a number of false starts, which delayed the adoption of preregistration (Dickersin & Rennie, 2003). But it is now possible to trace most research from inception through to completion, and it is clearer which research is funded by private donors who may have an interest in the outcomes of the research.

Drawing on the example of medicine, there is a broader movement toward preregistration of research that is just now entering the social sciences, including psychology, economics, and political science. Indeed, there is much optimism that adopting more stringent transparency standards should improve social science research (Humphreys et al., 2013; Miguel et al., 2014; Monogan, 2015). Registries for research designs have been established by the Evidence in Governance and Politics Network (EGAP), the BITSS, the American Economic Association's Randomized Control Trial (RCT) registry, the Registry for International Development Impact Evaluations (RIDIE) registry, and the Center for Open Science's Open Science Framework, among others. The expectation of these registries is as follows: preregistration creates the proper incentives to report (and publish) based on research design rather than the results, which makes it more likely that accurate causal effects—be they directional or null—come to light for the scientific community to be aware.

Preregistration admits a wide variety of possible designs. They could range from providing basic information about hypotheses and expected tests as with the basic EGAP preregistration option (http://egap.org/content/registration) to extraordinarily detailed analysis plans and mock reports as in many of the designs posted to EGAP (http://egap.org/design-registrations).

Although there is substantial theorizing about research transparency, there is very little empirical evaluation of actual preregistration practices. Some scholars have begun to register their designs, but very few of those designs have been published in political science (see Findley, Nielson, & Sharman, 2013, 2014, 2015; Gottlieb, 2016; Monogan, 2015 for additional examples). There are a few published examples in Economics and Psychology (e.g., Casey, Glennerster, & Miguel, 2012) in *Quarterly Journal of Economics* and the landmark preregistered observational study by Neumark (2001), but published results are not yet keeping pace with the growing interest in preregistration.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Preregistration complements replication policies by setting standards for earlier phases of the research, including the full set of data preparation and analysis operations to be conducted, which should reduce data fishing and in turn reduce publication bias. Although these incidents provide lessons about replication and data sharing, they are primarily aimed at catching publication bias caused by the researchers. They do little to address the role that the review process indirectly plays in the censorship of null findings

Our special issue received nine (out of 19) submissions for preregistered work that was yet to be fielded. As special issue editors, we were especially excited about submissions in which scholars submitted their plans prior to implementing their research. This allowed greater transparency in the research process and also enabled detailed feedback to the researcher before they went into the field.

Nevertheless, it is not clear that preregistration alone is a sufficient solution for publication bias. First, it is possible for authors to engage in hypothesis trolling, preregistering multiple measurements, hypotheses, and subsample analyses, thereby allowing themselves ample room to *p*-fish within their stated research plans. As multiple comparison corrections become increasingly common, these problems may not be as acute, because there is a penalty for each additional test conducted. In the most conservative case, for example, the Bonferroni correction requires a revised significant level cutoff at α/n where α is the standard significance level set by the researcher (typically .05) and n is the number of tests considered. Thus, as more outcome measures are considered, the significance level required goes to zero very quickly.

Second, as Simmons, Nelson, and Simonsohn (2011) point out, reviewers must be willing to take on the extra burden of downloading preanalysis plans and carefully determining whether the authors were faithful to their proposed design in addition to their other duties. If they do not, preregistration will not provide a reliable check on false positives. Even in the field of neurology, which has a much longer history of preregistration than the social sciences, follow-up studies have shown that 74% of work was never preregistered and the even work with preanalysis plans diverges considerably from the hypotheses and specifications of the preanalysis plan (Rayhill, Sharon, Burch, & Loder, 2015).

How Can Results-Free Peer Review Help?

In addition to considering a set of research designs, we explored an additional and complementary mechanism to address publication bias: results-free review. The special issue authors are in agreement about the merits of

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

replication, and three out of the four editors have preregistered their own research designs. We believe that results-free review has the ability to complement existing strategies for mitigating publication bias and increasing the transparency of the research process. What our special issue accomplished was reviewing all submissions—designs or completed studies—"results free."

Results-free review consists of authors submitting their manuscripts to the journal devoid of empirical results and then journals reviewing the manuscripts through to an accept–reject decision without ever seeing the results. One format is for a complete paper with all of the details of a normal submission with the exception of the empirical analysis and no mention of the actual results anywhere else in the manuscript. In another format, a manuscript is submitted that is close to a preanalysis plan, describing a study that has yet to be conducted. Both of these formats are "results free" in the sense that the results of the analysis are unknown to the reviewers and editor(s), but in the latter, the results are also unknown to the author(s).

How does results-free review help to solve the problem of publication bias? In short, reviewers assessed whether a theory was innovative, whether empirical tests were appropriate, and whether there were any obvious flaws in the design. If a research plan overcame all of these hurdles, it would be preaccepted for publication. As long as the researchers adhered to their plan, their work would be published regardless of the p values on their key causal variables. The idea behind this process is to encourage researchers to be more open and precise about their design on the front end by liberating them to be as open as possible about the fruits of their work on the back end.

This process can influence the decisions of both authors and reviewers. Authors had full knowledge that their work would be reviewed results free. As our focus was not on author incentives, we leave it up to the reader to provide conjectures on how this type of review shapes initial author decisions. Our focus in this special issue was to examine how reviewers evaluate papers without knowledge of the empirical results. This removes the bias of reviewers wanting to see work that is statistically significant or perhaps even counter-intuitive. In the next section, we discuss our exact process and then provide an evaluation of this process based on interpretation of the reviews.

Before doing so, it is important to emphasize that results-free review addresses only one set of problems that can lead to publication bias: professional incentives to produce significant findings that affect how authors analyze their data. We are skeptical that there is any kind of institutional design that will eliminate manipulation or fraud, and emphasize that norms of scholarly inquiry such as honesty, trust, and acceptance of fallibility are foundational to knowledge accumulation. However, results-free peer review can be helpful even in such a norm-based community, especially

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

when researchers' and reviewers' incentives lead them to privilege certain kinds of results over others.

Our Process

Our process began with a CFP in which we encouraged two types of submissions for results-free review. (See the appendix for the CFP.) This call was published on the *CPS* website, circulated through numerous email lists, and we wrote a short CFP for the *Washington Post's Monkey Cage*. We attempted to solicit manuscripts on all types of research, substantively and methodologically, that fit within the mandate of *CPS*. The *CPS* editors and special issue editors agreed that decisions on the manuscript would be made before seeing the final results, and the special issue editors would only check manuscripts to ensure they faithfully implemented the tests and analysis that were part of the final results-free submission.

In the first type of submission, we asked for a submission that approximated a preanalysis plan, instructing prospective authors that submissions for this special issue should provide designs that enable a reviewer to assess as fully as possible the theory, main hypotheses, design, feasibility, and potential contributions of the results. In the second type of submission, we invited submissions of otherwise complete manuscripts in which the results and discussion had been removed. For these submissions, the author(s) needed to provide a similar level of detail on the theory, design, *and* credible documentation that the results of the study were not posted or circulated in any way such that a peer reviewer could find and view the results and make a judgment on the paper with conclusions in mind. Preference was given to submissions that had not been previously reviewed at another journal. What united both types of submissions was that *reviewers could not use the results of the analysis to judge the value of the contribution*. The key difference between the two submissions was that the first type had not actually been carried out, whereas the second type had. In the end, this special issue features two articles where the data were only collected and analyzed after peer review (Bush, Erlich, Prather, & Zeira, 2016; Huff & Kruszewska, 2016), and one where the data were collected but the results unknown to the authors and reviewers (Hidalgo, Canello, & Lima-de-Oliveira, 2016).

As special issue editors, we were active in evaluating all manuscripts. Some manuscripts were judged not to fit the special issue, although we were open to any topic relevant to comparative politics. In most cases, these were relatively easy choices, but the harder decisions were about which manuscripts were of sufficient quality and provided enough detail to merit peer review. Given the novelty of this special issue, there seemed to be some

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

confusion about what constituted a results-free submission. Some submissions were very speculative and provided even less detail on the data collection and analysis plans than a research design section in a regular journal submission. Other submissions were on the topic of research transparency and did not conduct original research that fits within the mandate of *CPS*. The special issue editors read manuscripts and consulted with the standing *CPS* editors in a number of cases. In all, one manuscript was withdrawn by the author, eight manuscripts were desk rejected, and a final 10 manuscripts were sent out for peer review.

The special issue editors also helped with the selection of manuscript reviewers. Reviewers were largely chosen based on the substantive topic of the submitted manuscript, although some reviewers were selected based on their methodological expertise. Again, most of these decisions were easy, and for manuscripts, we had a long list of potential reviewers. Of the total of 43 reviewer requests we sent, 16 declined to review the manuscripts (37% turndown rate). This turndown rate is lower than the average *CPS* turndown rate of 47%.[10]

Reviewers submitted their comments through the regular *CPS* editorial mechanism, and the reviews were then sent to the special issue editors by the *CPS* standing editors. Both sets of editors jointly made the final decisions. Three of the 10 papers sent out for review were offered revise and resubmits. After revisions, all three papers were sent back to the original reviewers who all commented relatively positively, and the papers were then accepted for the special issue.

Once the decision to accept the manuscript had been made, that decision was the near-final decision on the manuscript, subject only to the constraint that the research was executed as planned. We instructed authors that deviations from the accepted research designs were acceptable, but had to be documented rigorously and discussed thoroughly. By asking that authors delineate the alterations made as a result of reviewer suggestions in the final article to clearly and publicly differentiate them from analyses that were preregistered, we gained novel insights into how the peer-review process shapes knowledge production and accumulation in comparative politics.

The Huff and Kruszewska piece is particularly enlightening in this regard. In the published article that follows, they present and interpret their results in line with their preanalysis plan. In addition, however, throughout the manuscript, they document how slightly different specifications from their preanalysis plans (i.e., specifying significance tests at 0.1 rather than 0.05 level or employing different baselines) would have altered their results. In discussing this presentational choice in their comments to the editors, the authors wrote:

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

> Presenting the results in this way is consistent with the goal of the special issue in promoting research transparency as it allows the maximum opportunity for the reader to draw their own conclusions from our results. Moreover, we think that doing so helps emphasize the importance of results-blind peer review in that it removes the incentive for authors to only present findings that are statistically significant while omitting models that are sensitive to design choices and outcome variable specifications.[11]

Potential Pitfalls and What We Learned

We began the pilot with optimism, yet we knew at the outset there were potential problems with the process. First, we were concerned that reviewers would be unwilling to review manuscripts without results, or that they might provide only cursory reviews not of the same quality as regular reviews. This was clearly not the case, where already burdened reviewers were willing to evaluate these manuscripts, and we were especially impressed with the quality of the reviews. The standing *CPS* editors agreed that these reviews were of higher quality than the average review.

A second concern is that these sorts of new forms of review can have implications for the types of authors willing to submit their work. For example, in a blog about preregistration, Joshua Tucker notes that untenured scholars may feel the most pressure to adhere to stronger norms of research transparency.[12] This could lead to imposing higher costs on more junior researchers, although we note that all of the authors in this special issue are junior scholars. Alternatively, we could observe faculty with tenure willing to embark on more "risky" forms of publication. In the case of this special issue, though, the review process generated submissions from all levels, ranging from graduate students to tenured faculty.

Third, results-free review, and especially preregistered designs, could actually lead scholars to invest *less* in theory development and select research questions that allow for hypotheses in different directions. In plain language, we worried that researchers would focus more on research projects where any empirical tests—positive, negative, or null—are interesting to readers. At the worst, this could lead to a type of hypothesis trolling where researchers propose a laundry list of hypothesis in a preregistration document, assuring themselves that there will be at least some significant results. We could have moved the discipline from data mining to hypothesis trolling.

Ironically, we are limited by research ethics and journal policy on how much of the insights we gained from the review process can be formally documented in this special issue on research transparency. Ideally, we could create an online archive of manuscript submissions, all reviews for the manuscripts, and include direct quotes from these reviews in this introduction.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Without going into detail, it is obvious that this leads to a number of ethical issues in that individual reviewers graciously provided reviews without the knowledge that these reviews would be quoted in this special issue to defend the claims of the special issue editors. As a compromise, we simply summarize reviewer comments in this special issue and provided detailed quotes from which we draw these summaries to the *CPS* editors. Thus, the full manuscripts, reviews, and the passages we are drawing upon have been verified by the *CPS* editors.

Our major concern turned out to have been largely unfounded; hypothesis trolling was specifically targeted and rejected by reviewers. Again, reviewer anonymity and author confidentiality prevent us from revealing specific comments, but reviewers noticed when, for example, manuscripts focused primarily on empirical data and proposed a wide range of theories and hypotheses to anticipate any and all findings. Such observations suggest that hypothesis trolling might be more common than we know in the work that is currently published. One reviewer was moved to comment to us that perhaps most manuscripts begin this way, with the theory being constructed post hoc and only then "sold" to the reader based on the results themselves. One advantage of results-free review over preregistration alone is that it nips this problem in the bud before the authors hit the jackpot on one of many hypotheses and rewrite the paper highlighting only the successful expectation and conclusion.

Our main findings from this exercise are in retrospect intuitive, but they were largely unanticipated. First, we found that reviewers placed a much *greater* focus on theory, the importance of the question, and most notably the relationship between theory and research design. This last point is worth emphasizing as some of our submissions had important theoretical contributions and rigorous research designs, but reviewers consistently commented on weak links between theory and analysis.

Relatedly, reviewers in our pilot insisted on a great deal of country context and knowledge to understand the design choices, adjudicate their importance, and think about external validity. The combination of designs focusing on causal inference and results-free review appeared to emphasize the importance of area-specific knowledge.

Third, reviewers (and the special issue editors) struggled to identify the criteria for which studies would be publishable even with null findings. Which null results are valuable and which can be dismissed due to research design issues? Although null findings have given considerable discomfort to scholars in the social sciences, relatively little discussion exists on how null findings should be treated in the review and publication process.

Finally, we did not receive a single qualitative submission. We attempted to reach out to qualitative researchers through explicitly qualitative research

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

channels, and were hopeful that we would receive at least some non-quantitative papers to review. Interestingly, our first finding that authors and reviewers valued substantive importance and theory could very well have privileged qualitative work. Alas, we had no submissions of this type and we speculate below as to the causes of this bias.

We flesh out the discussion of each of these issues below

Theory and Substance

The independent evaluations of the four special issue editors were in complete agreement regarding the rigor and focus of the reviews. All four of us were struck by the reviewers' extensive focus on each manuscript's theory and substance. The reviews were in comparable length to a regular journal review but did not have the same focus on the interpretation of results. Reviewers obviously made comments on the methodology, control variables, and issues with the empirical research design. But we judged these reviews as focusing much more on the "substance" of the manuscript and the relationship between the question, the theory, research design, and the potential contribution.

We believe that this outcome could very well be the greatest success of the special issue. Experimentalists, who focus intently on the identification of causal effects, have been a key group pressing for greater transparency, including preregistration and results-free review. And yet scholars point out that theory may be left behind in the race toward better and better identification. John Huber (2013) lamented that a laser like focus on causal identification in research designs might lead scholars to eschew difficult social science questions in favor of queries that allowed for designs more closely approximating randomization. Huber was making a nuanced point, but the article triggered broader water cooler discussions about whether well-identified work was also theoretically grounded. And David Laitin (2013) articulated a related concern that preregistration might undermine the productive feedback loop between empirical research and theoretical exploration. These concerns may be warranted, but the results of this exercise demonstrate that theory need not be lost; indeed, given the right peer-review incentives, theory and substance may carry the day.

Of course, political scientists may disagree on what should be emphasized during the review process, but our special issue clearly demonstrated that this process shifted the focus of these manuscripts toward the substance. By far, the most common concern from our pool of reviewers was inattention to theory. Of course, we observed the common gripes about lack of acknowledgment of the extant literature and limited engagement with major

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

contributions. More poignantly, however, what became clear is that it was impossible for a research design to be atheoretical and survive results-free review. Every stage of the enterprise from choice of location to operationalization to specification to analysis of heterogeneous effects depends on well-defined theory as the guiding light. Again, anonymity and confidentiality prevent us from providing examples, but time and again, imprecise theory made it impossible for reviewers to determine whether the research designs could help answer the author's ultimate question.

Reviewers were also quick to note when the theory section seemed off the mark, responding to the wrong literature and missing critical antecedents that would make it more broadly appealing. Whether or not journals should implement this process, either for all reviewers or a subset of reviewers for each manuscript, is a question that we cannot answer. But, in our experience as authors and reviewers, the real effort to unpack the theoretical questions as a way of understanding the research design was quite different from what we have seen in our experience writing and reviewing otherwise. In our experience, when results are available, the discussion between authors and reviewers becomes one of "what theory are these results consistent with?" When results are not available, then the theory has to stand on its own. We now conjecture that results-free review could reinforce a more productive interplay between theory and empirics.

The Return of Area Expertise

For decades, there has been a deep tension between students of comparative politics and regional specialists (see, for example, Bates, 1996). Traditional area specialists have criticized mainstream comparative politics, especially large-N cross-national work, as devoid of local context. Many cross-national comparative politics scholars have flipped the criticism on its head, claiming that country experts' work is of little utility beyond their very small and tightly knit community (Pepinsky, 2015).

As with the critical importance of theory, one key lesson of our pilot is that results-free review of field research rewarded greater emphasis on areas studies knowledge as essential for building more compelling research designs. In numerous reviews, referees demanded greater local specificity to understand the implications, internal validity, and generalizability of the design and predicted results. In 100% of the submissions that had a field-based component, the question of whether the authors had the adequate area expertise to carry out and make sense of the research results came up.

This happened in a number of different ways. Some reviewers wondered whether treatments would be effective in a particular country context given

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

reviewers' specific knowledge of how institutions work there. Others asked about the meaning of key variables in particular national contexts. Others focused on external validity and the broader theoretical impact of findings from a particular country, given the unique climate for an experiment there. These questions even came up in two of the successful manuscripts, forcing the authors to better defend their choices and offer more thorough descriptions of context.

As with theory, it appears to us that reviewers were liberated to challenge researchers on these fundamental questions, because they did not have to deal with the distraction of the empirical results. In the three successful submissions, such focused attention on context and local knowledge led to what we perceive as major improvements in the authors' research designs. Authors were forced to address local nuances that might affect interpretation and choose designs that would best help readers think about generalizing to other contexts. For proponents of a new synthesis of area studies and comparative politics, one that eschews the battles between local context and general social science (see Malesky, 2008; Pepinsky, 2015), this is an encouraging result.

Null Results

The third result of our pilot is so provocative it divides this special issue team. As noted above, numerous reviewers expressed frustration in reviewing work without results, in some cases admitting their own biases, and in other cases making clear that the direction and size of the results are a core part of the intellectual contribution. There are two interrelated problems that the subject of null findings poses for review. The first has to do with acclimating to a new way of thinking about null findings—that they may be meaningful theoretically. The second is the question of what types of null findings are worthy of publication.

It seems especially difficult for referees and authors alike to accept that null findings might mean that a theory has been proved to be unhelpful for explaining some phenomenon,[13] as opposed to being the result of mechanical problems with how the hypothesis was tested (low power, poor measures, etc.). Making this distinction, of course, is exactly the main benefit of results-free peer review. Perhaps the single most compelling argument in favor of results-free peer review is that it allows for findings of non-relationships. Yet, our reviewers pushed back against making such calls. They appeared reluctant to endorse manuscripts in which null findings were possible, or if so, to interpret those null results as evidence against the existence of a hypothesized relationship. For some reviewers, this was a source of some consternation: Reviewing manuscripts without results made them aware of how they were

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

making decisions based on the strength of findings, and also how much easier it was to feel "excited" by strong findings

This question even led to debate among the special issue editors on what are the standards for publishing a null finding? For example, let us return to the LaCour and Green (2014) paper once again. Imagine that this research was faithfully conducted and submitted without results. Would this paper merit publication in a prominent journal? If our expectation was a null or small impact based on substantial prior research indicating just that, would the study be worth publishing? If we knew there was a large finding, would that change our evaluation of this paper? Again and again, reviewers posed some version of the question: If the tested hypotheses proved insignificant, would that move debates in this subliterature forward in any way? In many of the rejected papers and even one of the accepted papers, the answer was no.

There were three reasons that reviewers reached this conclusion. First, a null finding would not be interesting because the reviewer found the theory to be implausible in the first place. Proving that the implausible was in fact implausible is not a recipe for scintillating scholarship.

The second was a variant of Occam's razor. Reviewers did not believe that the author had adequately accounted for the simpler, alternative theory to explain the underlying puzzle that motivated their research. In this instance, a null result would only reinforce the notion that the more parsimonious theory was superior, or that a natural experiment was confounded by unobservable selection.

Third, there was too much distance between the articulated theory and the abstract field, lab-in-field, or survey experiment articulated in the paper. The theory invoked a compelling concept, but the proposed research design failed to adequately capture it or stretched the meaning of the concept to the point of unrecognizability. In this case, a null result would only prove the empirical test was inadequate for the bigger question. This was a common criticism of experimental research.

None of these dismissals of proposed research plans are new problems or unique to results-free review. They are a standard part of the way scholars evaluate research. The interesting implications for results-free review manifest themselves in how strategic authors may alter their research agenda to survive the review process. Knowing that they have to convince a skeptical reviewer that a null finding is interesting, they may choose to abjure big questions and paradigmatic shifting scholarship for incremental research designs. Remember also that a laundry list of hypotheses and potential heterogeneous effects will not suffice either. Our reviewers were quick to spot and reject this type of hypothesis trolling.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Three author strategies would seem most plausible. First, authors place themselves between two competing theories with contrasting observable implications, posing their research design as the distinguishing test. For example, does fiscal decentralization decrease or increase corruption? Here, a null finding might rule out one of the competing hypotheses.

Second, authors may offer their research design as the first or a better test of prevailing theory or logic that has been inadequately tested in the literature. The theory of deliberative democracy, for instance, offers a number of very clear implications about how deliberation should affect the thinking and behavior of citizens, yet, most of these have been subjected to only limited empirical testing. If designed properly, this would be interesting purely because the potential target would be well known. Again, reviewers reacted quite negatively to this type of approach. Most referees wanted authors to build on the existing literature in important ways or to thoroughly explain why the observational work of previous generations was flawed.

Finally, authors might offer a test of a hypothesis that is the next logical step within a prevailing and well-traveled research paradigm. In the American politics literature, theories regarding voter mobilization efforts and turnout are the closest to the type of incremental progress we have in mind.

All of these strategies would likely fare better in results-free review than a brand new theory, built directly from first principles, or paradigm-shifting theory that challenges the prevailing wisdom in the literature. However, all three approaches are predominantly empirical, building upon existing theory, rather than creating it. In Thomas Kuhn's terminology, results-free review would engender a lot more normal science.

And here is where the disagreement among the co-editors is most severe. Some of the special issue editors applaud this potential trajectory, arguing that it is time that political science de-emphasized grand theorizing, focusing on a gradual accumulation of knowledge that specifically includes a large catalog of theories that have not proved useful. These editors argue that there will always be outlets for big think pieces, but there is still not enough room for the hard, plodding empirical confirmation of the discipline's theorists.

The other co-editors worry about the damaging result this trajectory would have on creative scholarship. They worry that there are still big questions out there to be asked. In fact, as recent events have shown, on some of the most vital questions to mankind such as economic inequality, international immigration patterns, the role of aid in disaster relief, and the resilience of state institutions to global pandemics, the depth of political science scholarship has proved wholly inadequate to society's needs. This is not the time, they argue, to discourage big theory and narrow the lens of the field's most ambitious scholars.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

There is one alternative that we have not discussed that may provide a way around the problem of what to do with null results. That is for authors and reviewers alike to abandon null significance hypothesis testing altogether. The conceptual problems with null significance hypothesis testing should be well known to political scientists (e.g., Gill, 1999), but periodic calls for a Bayesian alternative have yet to unsettle long-established practice. In a blog post reflecting on problems of *p*-fishing and experiments, Simon Jackman commented, "From the Bayesian perspective, all this stuff is kind of ridiculously overblown, a consequence of an unthinking acceptance of $p < .05$ as a model for scientific decision making, point null hypothesis testing, the whole box and dice."[14] But the problems are deeper. Jackman invokes Berger and Sellke (1987), who demonstrate that small *p* values do not (necessarily) correspond to strong evidence against a null hypothesis that a parameter is zero. And as is well known, even in the standard frequentist setting, large *p* values are not evidence that a parameter *is* zero.

If reviewers and authors did not attribute substantive meaning to tests of statistical significance then there would be no statistical significance filter. What would replace null significance hypothesis testing remains unknown. But we emphasize that authors and reviewers used statistical significance as a shorthand for adjudicating whether effects exist or not. This indicates to us that getting away from the very premise that there is such as thing as "null" results (to say nothing of "non-results") will require a significant departure from current practice. Perhaps one result of our pilot study is to highlight not just the practical difficulties that reviewers face with null results, but the conceptual and theoretical problems with null results that extend to the vast majority of published research in political science.

Method

A final observation is that our special issue generated a very particular type of submission. The vast majority of submissions that we received were for survey or field experiments, and the remainder involved the statistical analysis of quantitative data. We received no submissions of qualitative case studies, historical comparisons, or ethnographic research.[15]

Why would this be? It is not possible to answer this question definitively based on our own experiences, but there are at least four possibilities. One is that our CFP happened to have been read primarily by people working in the new experimental tradition in political science. If so, and despite our efforts to the contrary, we simply failed to reach out broadly enough to include a representative sample of research in contemporary political science.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Another is that authors using different kinds of methodologies saw our announcement, but believed that we were looking primarily for experimental, or at least statistical, research. We did not intend to elicit only statistical or experimental submissions, but we also did not take extra steps to encourage specifically qualitative methods in our submissions. Although we did attempt to reach out to a group that coalesces around the study and practice of qualitative methods, the effort appears not to have been sufficient. This would be our own failure and not a limitation of results-free review.

A third reason why we did not receive qualitative submissions may stem from the reputation of *CPS* as a quantitative journal. The journal maintains no explicit policy about methodology, and is working to change the reputation, but it nonetheless is still largely known as the central quantitative journal in comparative politics. As qualitative researchers are well represented within comparative politics, a large number of possible submitters may have been influenced by the journal's reputation.

Still another is that qualitative case studies, comparative historical analyses, and other similar types of research *cannot be preregistered* and that *results cannot be removed from case studies*. These are types of research in which scholars generally accept that theories and arguments are informed by the interaction between a researcher's initial hypotheses—in some cases little more than hunches—and the specifics of a case. We believe this type of work is a valuable contribution to political science scholarship, but we can imagine the complexity of submitting this work preregistered and/or results free.

The core principle of preregistration, that hypotheses must be specified before the researcher collects and analyzes the data, is simply incompatible with approaches that prioritize the reciprocal engagement between theory and evidence. This seems especially difficult for qualitative and historical types of research. More fundamental to our special issue, where half of the manuscripts were not preregistered, all of the papers were submitted results free. The premise of results-free peer review—that it is possible to describe a research enterprise without reference to the data it produces—is inconsistent with the way that we actually conduct qualitative comparative analyses and in-depth case studies (see, for example, Yom, 2015). Importantly, our argument is not that such research is unscientific—it certainly can be, and in fact, such research can fit well within a positivist epistemology. The point is simply that inductive research, and various kinds of mechanismic-centered qualitative and historical research, cannot be described without reference to the data.

The special issue editors all have different kinds of methodological expertise, but for those of us who have worked with historical and archival materials, and who have paired these with in-person interviews with important policy makers, the tensions between preregistration and results-free peer

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

review, on one hand, and careful qualitative research, on the other, seem insurmountable. Although we have no reason to conclude that results-free peer review must prevent such research, such that shared standards can never be developed, results-free peer review would likely have serious implications for current practice.

Specifically, we suggest that results-free peer review has an affinity for a normal science view of social scientific research. Results-free peer review is most feasible when authors are working within established research traditions that work with clear and long-established hypotheses. It also has an affinity for research that uses formal analytical tools to generate deductive hypotheses. In both of these cases, a positivist epistemology undergirds the research enterprise. It is for this reason, we suggest, that experimental methods were particularly attractive to authors who submitted manuscripts to *CPS*; they themselves have a natural affinity for estimating causal effects from designs drawn from well-established theory.

By contrast, it is difficult to see how interpretivist and other post- or non-positivist epistemologies would work with results-free review. A core feature of interpretivism is the rejection of any strict distinction between theory and data, so that the struggle for many interpretivists is to leave aside theories and assumptions about the social world. Especially in ethnographic or hermeneutic research, the research enterprise seeks to uncover how meaning is made, or to come to understand the lived experiences of interlocutors or the texts that they have produced. Only through research itself do these meanings become clear. It is certainly possible to plan an ethnographic project, or to list a series of texts or archives one plans to consult, but it makes little sense to list hypotheses and the data to be collected to test them, because neither the hypotheses nor the data can be known in advance.

Case study research in the standard positivist mode lies in between these two extremes. We find it useful to draw on Lieberman's (2005) distinction between "model-building" and "model-testing" small-n analyses. The former is characterized by much less certainty about the theory or the data—in Lieberman's words, "the scholar engaged in [model-building small n analysis] does *not* proceed with the notion that a fully specified model is available and must develop explanations for the puzzle of varied outcomes" (pp. 443). This type of inductive, exploratory research fits less obviously with a results-free model of peer review—even when the research is complete and has been written up—because describing the qualitative data being used may itself be part of the process that generates the new theoretical insight. By contrast, model-testing small-n analysis that draws on cross-case statistical findings to justify intensive study of whether particular cases are consistent with those findings fits more naturally in the experimental or observational templates

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

described above. Although we did not encounter any such submissions, we find it easier to conceive of results-free description of such a case study or historical analysis.

Future research by qualitative methodologists into the possibility of preregistering historical, ethnographic, or otherwise exploratory research may help to tell us whether the very nature of our special issue itself discouraged qualitative submissions. But we see here a parallel with the tensions that we identified in the previous section on null results. Some political scientists may welcome preregistration precisely because it places greater emphasis on political science as a normal science. Others will see that as a substantial drawback.

It is hard to escape the conclusion, though, that any requirement that research manuscripts have been preregistered will almost certainly affect the types of submissions that a journal receives. One possible consequence is a bifurcation of publication outlets, and as a result, of researchers. One set of researchers adheres strictly to a normal science template to produce manuscripts that are eligible for journals that insist on results-free review, while others adhere to and are assessed on a very different set of standards in a different set of journals. For the discipline as a whole, this would almost certainly generate divisions and inequalities.

An Overview of the Articles

Of the 19 articles originally submitted, and then the 10 sent out for review, three submissions were granted revise and resubmit status. The authors of the three papers made revisions to the designs or results-free submissions and then resubmitted, and each was subsequently conditionally accepted for publication. The papers were accepted conditional only on finalizing the papers without any deviations so large as to be out of the spirit of what was reviewed by the referees. Otherwise, regardless of whether results came back weak or strong substantively, be they null, positive, or negative, the papers would still be published. We emphasize that the submissions were conditionally accepted prior to any results being available to the standing CPS editors, to the special issue editors, and to all of the anonymous reviewers. The three accepted papers appearing in this volume are Bush et al. (2016) "The Effects of Authoritarian Iconography: An Experimental Test," Hidalgo et al. (2016) "Can Politicians Police Themselves? Natural Experimental Evidence from Brazil's Audit Courts," and Huff and Kruszewska (2016) "Banners, Barricades, and Bombs: The Tactical Choices of Social Movements and Public Opinion."

In their paper, Bush et al. experimentally examine the effects of authoritarian images, or iconography, on citizen behavioral compliance and regime

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

support. Tying into the broader literature on authoritarian survival and cults of personality, the authors contend that the use of iconography—posters, sculptures, seals, or insignias bearing images of authoritarian leaders—should work through mechanisms of legitimacy, self-interest, and coercion to generate greater compliance and regime support among citizens. Bush et al. carried out their laboratory experiment in the United Arab Emirates, a rising regional, authoritarian power. The authors submitted a research design for the special issue and only carried out the actual experiment once the review process was complete and they received a conditional acceptance. They find no evidence that iconography affects support for the Emirati regime in their experimental analysis. The authors probe the null result with additional tests, finding that standard explanations such as insufficient power or measurement error are unlikely explanations for the lack of significance. As a result, this article is an excellent example of a research project demonstrating that existing theoretical expectations may have been overblown.

Hidalgo, Canello, & Lima-de-Oliveira examine horizontal accountability mechanisms for government political institutions. They explore how differential assignment to public audit courts affects punishment of lawmakers explicitly leveraging a natural experiment whereby government agencies and subnational governments are assigned by lottery to municipality-level councilors. This paper was submitted after data had been collected but before it had been analyzed. In lieu of results, the authors provided a detailed preanalysis plan identifying how each hypothesis would be tested so that the reviewers and editors would have a comprehensive picture of all procedures. Once conditionally accepted, the authors then added the results, showing that when institutions shield auditors from political interference, they punish lawbreaking politicians more than do auditors who do not enjoy such insulation.

The Huff and Kruszewska study engages the broad contentious politics literature with an examination of how the degree of extremeness of tactical choice influences public opinion about a movement. Although this central question has long been a concern among scholars, identifying the effects of tactical choice is an exercise laden with pitfalls. The authors use a novel survey experiment design in Poland that systematically varies tactical choice and then considers the extent to which subjects believe that the government should negotiate with a movement as well as how many concessions should be offered. The authors' submission was a research design that had yet to be implemented and the authors updated the design in response to the reviewers' and editors' feedback. The authors set up a theoretical horse race among three approaches: the benefits of extremism, the benefits of moderation, and the no-concessions policy. Supplementing standard statistical techniques with a

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

structural topic model that allowed for text analysis, the results are broadly consistent with the no-concessions theoretical approach.

Final Thoughts

In many ways, this special issue exceeded our expectations. First, *CPS* will publish three excellent studies that we believe will have an impact on the discipline. This is obviously first and foremost due to the authors' hard work, but we also credit the reviewers for their contributions. Our subjective evaluation was that the reviewers provided extremely high-quality reviews and most likely set the bar for these papers higher than for the normal submission. Whether or not raising this bar is advisable is debatable, but we can clearly state that this form of review lead to papers that were of the highest quality. We would love to see a top journal adopt results-free review as a policy, at very least allowing results-free review as one among several standard submission options. In thinking about whether the pilot could be applied more generally, and reflecting on some of the logistical issues that we faced in this process, three practical considerations arose.

First, this form of review could lead to new incentive problems. The least serious of which is that journals that accept results-free papers might be flooded with null results as academics open up their file drawers. This could be a problem for an individual journal where this form of review leads it to specialize in null results, and that reviewers, knowing it is results-free, have priors that the submitted paper probably has weak or null results. Although this may lead to some immediate problems as journal pipelines get flooded, overall it should correct for the biased distribution in p values and open up outlets for high-quality papers with null findings. We note that for the three papers published here, that one of them had null results as the main finding, although this result was not known to the authors at the time of submission.

Second, as a discipline, we might want to rethink when it is appropriate to present and circulate work. The requirement that papers are not fielded or have not been presented put the work submitted to the special issue at an enormous disadvantage. Polished working papers frequently have benefited from insights of discussants and participants, are often sent to peer-reviewed journals and are rejected by the first journal, (hopefully) revised based on the reviewer criticisms, and then submitted to another journal. The submissions to this special issue did not benefit from this same cycle of presentation, feedback, and revision, leading at least some initial submissions to be more "green" than what we imagine is normally submitted to *CPS*. The trade-off of an author hiding results from reviewers is that they may also hide the work from the research community, limiting the amount of feedback they receive,

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

and possibly affecting the amount of scholars who know about this work even after publication. Currently, many papers become well known even before they are published, which may or may not be desirable, but that might not continue under a largely results-free review standard. This concern is not as serious for preregistered designs that would be reviewed results free—in those cases, one could imagine scholars circulating the designs widely with no fear of results getting out. To the extent that designs were to be accepted ahead of conducting the research, this could mitigate the problem of publication bias as well as under developing papers. However, for other types of research, there is a clear trade-off, and results-free review will not work if authors are forced to strip out results from papers that have already been circulated.

Third, for preregistered studies, a results-free review mechanism could add yet another bottleneck to an already long research process. Field researchers often have to develop a project idea, conduct a pilot study, secure funding, go to the field and implement the research, and write up the results. And this sentence even sugarcoats the difficult road all researchers travel in the timing of research projects, where formal deadlines (grant applications, tenure clocks) all to some extent dictate the timing of projects, the academic calendar and teaching obligations shape when time is available, and personal factors, such as the timing of breaks in child care, require a Tetris style mastery of scheduling to pull off successfully.

Field researchers submitting preregistered designs, if they followed our framework, must now wait for peer reviews and editorial decisions before they can move onto the next step. For researchers who have not fielded a study, they must wait for suggestions that may completely reshape their proposed study. For researchers who have already fielded the study, they may be waiting for reviewers before they even begin a preliminary analysis of the data. If the author receives a rejection from the journal, she or he may consider sending their proposal to another journal, further pushing back the fielding of their project. This process could add years to the time between idea generation and the beginning of the field research. So it is important to note that what we perceived as the gold standard of transparency (submission of the design prior to fielding the survey) imposed clear costs on the researcher.

On the flip side, there may be some benefits only possible with this style of review. For one, currently when a scholar carries out a research project, there is often a very long delay before those findings get published. Some may argue, drawing on ethical principles, that policy-relevant research should be made available as soon as possible. If the review process is fully complete before the research is executed, then the study can be put into the publication pipeline almost immediately following the time that researchers carry out the

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

study, substantially increasing the timeliness of research for the political problems under study. Moreover, if a scholar's design is already accepted for publication, they may find it easier to secure funding if they have not already done so. And the difficulty of carrying out the research may be lessened if organizational partners and other stakeholders are aware that the research will not only eventually see the light of day but will do so in the near future.

These are practical considerations. But moving forward, our reflections in this essay do raise some consistent themes.

1. Is a full recording of all steps of a research project, from conceptualization to empirical testing, actually possible? What is the limit to the information that we wish to record, and how much of the intellectual architecture should peer reviewers have access to when evaluating a manuscript?
2. How should we interpret null results? Why are authors and reviewers alike so willing to accept the null hypothesis significance testing paradigm, yet reluctant to conclude that insignificant results are evidence against particular hypotheses? Might a Bayesian framework provide an alternative foundation for hypothesis testing, one that puts more structure on hypothesis testing as a decision problem rather than declaring results to be significant or not?
3. To what extent is the affinity of results-free peer review with a normal science view of comparative politics inevitable? Can inductive and exploratory research fit within such a research paradigm?
4. Is political science mature enough of a discipline that researchers should only ask questions in which any results, null or otherwise, are interesting? Does this vary by subfield, topic area, or research question?

Both supporters and critics of results-free peer review will benefit from keeping these in mind.

Appendix

Research Transparency in the Social Sciences

Issue editors: Michael G. Findley, Nathan M. Jensen, Edmund J. Malesky, and Thomas B. Pepinsky.

We invite proposals for a special issue of *Comparative Political Studies* (*CPS*) on research transparency in the social sciences. Proposals for original research papers using quantitative or qualitative approaches, and collecting

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

quantitative or qualitative data are all encouraged. The deadline for submitted proposals is October 15, 2014, for a Special *CPS* issue scheduled to appear in 2015-2016 academic year.

There is growing momentum in the natural and social sciences for greater transparency in research. For example, see the Evidence in Governance and Politics Network (EGAP; www.egap.org) and the Berkeley Initiative for Transparency in the Social Sciences (BITSS; www.bitss.org). Although there are varied objectives driving the shift toward greater transparency, one of the key motivations is to avoid publication bias, the result of peer-review processes that privilege the significance of results over the theoretical contribution or integrity of the research design. On the contrary, critics of preregistration argue that it can handcuff authors, leading to journals filled with projects that are less theoretically innovative and path breaking than would otherwise be possible.

This Special Issue of *CPS* will help to assess the potential benefits and costs associated with new models of the publication process by studying how new models can work in practice. Transparency should obviously be a central objective in contemporary social science, but what are the costs? Do strict preregistration protocols commit scholars to carry out projects that are unfeasible, or dissuade creative dialogue between theory and data? Is it possible for a full recording of all steps of a research project, from conceptualization to empirical testing? How will manuscript referees respond to manuscripts without results or conclusions? These questions cannot be settled in the abstract.

The Special Issue aims to study the role of full transparency in research in two ways: (a) accepting work based on prospective research designs, and (b) opening up field notes and last-minute alterations in the research design through online archiving (as with replication data). Articles in the Special Issue will be bookended by two articles by the editors, which introduce the goals of the project and critically evaluate the pros and cons of preregistration and research transparency in political science.

To this end, we invite one of two types of submissions:

1. Full research designs for *prospective* research projects that have not yet been conducted.
2. Full research designs for projects that have already been conducted, and for which any discussion of results has been stripped out of the manuscript.

If the first type of submission, the design needs to be a thorough project prospectus, sometimes referred to as a preanalysis plan. Although there are

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

multiple ways to construct a preanalysis plan, submissions for this special issue should provide designs that enable a reviewer to assess as fully as possible the theory, main hypotheses, design, feasibility, and potential contributions of the results. This information should be sufficient to allow reviewers to reach a firm conclusion on the project, and ultimately accept or reject the project for publication in the special issue.

If the second type of submission, the author(s) need to provide a similar level of detail on the theory, design, *and* credible documentation that the results of the study are not posted or circulated in any way such that a peer reviewer could view the results and make a judgment on the paper with conclusions in mind. Preference will be given to submissions that have not been previously reviewed at another journal.

Once the designs have been submitted, they will be sent out for full peer review. Designs will be accepted, rejected, or invited to make revisions with resubmission. Once a determination has been made on the design, that decision will be the near-final decision on the manuscript, subject only to the constraint that the research is executed. Deviations from the accepted research designs are acceptable, but need to be documented rigorously and discussed thoroughly. In fact, it is expected that authors of projects that have already been conducted will be asked by reviewers to perform analyses outside of their initial research protocol. This is a normal part of the peer-review process: We ask that authors delineate the alterations made as a result of reviewer suggestions in the final article to clearly and publicly differentiate them from analyses that were preregistered. This will provide the editors with unique insight into how the peer-review process shapes scientific knowledge and accumulation.

Authors of research papers that are invited to move forward with publication will need to make available all background documents including coding notes, full replication files, and so on. To facilitate this process, authors will be eligible for a US$5,000 grant provided through the University of Texas at Austin to offset the costs of gathering and making available the required documents, notes, and so on.

As with regular submissions, the *CPS* permanent editors will make a definitive acceptance or rejection based on how authors address the reviewers' comments, but will not make an independent evaluation of the paper based on the final results.

Proposals should follow the standard *CPS* submission requirements for normal articles, but should be submitted to directly to the special issue editors at transparency@ipdutexas.org. Please indicate "*CPS* Special Issue Submission" in the subject line. We encourage you to contact the special issue editors if you have any questions at the above email.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Acknowledgments

We thank the standing *Comparative Political Studies* (*CPS*) editors, David Samuels and Ben Ansell, for supporting this new and unique special issue on preregistration and results-free review. As this had never been tried in political science, we recognize they put in a great deal of effort accommodating the new challenges of this review process.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Study registries are also valuable to help document research projects that did not result in publication. These design registries can be valuable in documenting which studies do not ultimately get written up as part of a research project. In this introduction, we do not focus on this important question and rather address how registration or results-free review affects the evaluation of manuscripts.
2. See a recent discussion in the *Economist* for some examples: "Trouble at the Lab" (2013).
3. We discuss this in more detail below.
4. A follow-up study by Broockman and Kalla (2016) finds large canvasing effects but this is unrelated to the gender identity of the canvaser.
5. Six common *p*-hacking tactics are as follows: (a) stopping data collection once $p < .05$; (b) analyzing many measures but report only those with $p < .05$; (c) using many specification but only report those with $p < .05$; (d) using covariates to get $p < .05$; (e) excluding participants to get to $p < .05$; and (f) transforming the data to get $p < .05$ (Simonsohn, Nelson, & Simmons, 2014).
6. There have been research fraud scandals in several disciplines, but one of the most public involving political science is LaCour and Green (2014); see above for discussion.
7. The issue of significance versus insignificance should also be separated from the issue of whether a result is close to or far from zero. On this point, see Hartman and Hidalgo (2015).
8. For an excellent overview of this controversy, see Cassidy (2013).
9. At the time of writing this paper, the "first place" author had 183 retractions. See http://retractionwatch.com/the-retraction-watch-leaderboard/
10. We thank the *Comparative Political Studies* (*CPS*) editors for providing this data.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

11. Correspondence with editors (July 28, 2015).
12. See http://blog.oup.com/2014/09/pro-con-research-preregistration/
13. We do not use the language here of a theory being "wrong" or "false," because theories are neither true nor false (Clarke & Primo, 2012).
14. The post is available here: http://webcache.googleusercontent.com/search?q=cache:DVdwW5x20DcJ:jackman.stanford.edu/blog/%3Fp%3D2708&hl=en&gl=us&strip=1&vwsrc=0
15. We did receive a few proposals for theoretical analyses of important concepts in political science. These were uniformly of a lower quality than our other submissions, and were from scholars who had not yet received a PhD, so we hesitate to conclude much about them.

References

Bates, R. H. (1996). Letter from the president: Area studies and the discipline. *American Political Science Association–Comparative Politics*, *7*(1), 1-2.

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American Statistical Association*, *82*, 112-122.

Broockman, D. E., & Kalla, J. L. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*, 220-224.

Broockman, D. E., Kalla, J. L., & Aronow, P. (2014). *Irregularities in LaCour (2014)*. Unpublished manuscript, University of California, Berkeley.

Bush, S., Erlich, A., Prather, L., & Zeira, Y. (2016). The effects of authoritarian iconography: An experimental test. *Comparative Political Studies*, *49*, 1704-1738.

Carey, B. (2011, November). Fraud case seen as a red flag for psychology research. *The New York Times*. Retrieved from http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html?_r=1

Casey, K., Glennerster, R., & Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *Quarterly Journal of Economics*, *127*, 1755-1812.

Cassidy, J. (2013, April). The Reinhart and Rogoff controversy: A summing up. *The New Yorker*. Retrieved from http://www.newyorker.com/rational-irrationality/the-reinhart-and-rogoff-controversy-a-summing-up

Clarke, K., & Primo, D. (2012). *A model discipline: Political science and the logic of representations*. Oxford, UK: Oxford University Press.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., …Van Der Weyden, M. B. (2004). Clinical trial registration: A statement from the international committee of medical journal editors. *The New England Journal of Medicine*, 351, 1250-1251.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, *263*, 1385-1389.

Dickersin, K., & Rennie, D. (2003). Registering clinical trials. *Journal of the American Medical Association*, *290*, 516-523.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Findley, M. G., Nielson, D. L., & Sharman, J. C. (2013). Using field experiments in international relations: A randomized study of anonymous incorporation. *International Organization*, *67*, 657-693.

Findley, M. G., Nielson, D. L., & Sharman, J. C. (2014). *Global shell games: Experiments in transnational relations, crime, and terrorism*. Cambridge, UK: Cambridge University Press.

Findley, M. G., Nielson, D. L., & Sharman, J. C. (2015). Causes of non-compliance with international law: Evidence from a field experiment on financial transparency. *American Journal of Political Science*, *59*, 146-161.

Finkel, S. E., Pérez-Liñán, A., & Seligson, M. A. (2007). The effects of U.S. foreign assistance on democracy building, 1990-2003. *World Politics*, *59*, 404-439.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*, 1502-1505.

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*, 641-651.

Gerber, A., & Malhotra, N. (2008). Do statistical reporting standards affect what is published: Publication bias in two leading political science journals. *Quarterly Journal of Political Science*, *3*, 313-326.

Gerber, A., Malhotra, N., Dowling, C., & Doherty, D. (2010). Publication bias in two political behavior literatures. *American Politics Research*, *38*, 591-613.

Gill, J. (1999). The insignificance of null hypothesis significance testing. *Political Research Quarterly*, *52*, 647-674.

Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 945-1017). New York: Elsevier.

Gottlieb, J. (2016). Greater expectations? A field experiment to improve accountability in Mali. *American Journal of Political Science*, *60*, 143-157.

Hartman, E., & Hidalgo, F. D. (2015). *What's the alternative? An equivalence approach to balance and placebo tests*. Unpublished manuscript, Princeton University, NJ.

Herndon, T., Ash, M., & Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, *38*, 257-279.

Hidalgo, F. D., Canello, J., & Lima-de-Oliveira, R. (2016). Can politicians police themselves? Natural experimental evidence from Brazil's audit courts. *Comparative Political Studies*, *49*, 1739-1773.

Huber, J. (2013, June). Is theory getting lost in the "identification revolution"? *The Political Economist*, pp. 1-3.

Huff, C., & Kruszewska, D. (2016). Banners, barricades, and bombs: The tactical choices of social movements and public opinion. *Comparative Political Studies*, *49*, 1774-1808.

Humphreys, M., de, la, Sierra, R. S., & van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, *21*, 1-20.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Ioannidis, J. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Journal of the American Medical Association*, *279*, 281-286.

LaCour, M., & Green, D. (2014). When contact changes minds: An experiment on transmission of support for gay equality. *Science*, *346*, 1366-1369.

Laitin, D. (2013). Fisheries management. *Political Analysis*, *21*, 42-47.

Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review*, *99*, 435-452.

Malesky, E. J. (2008). Battling onward: The debate over field research in developmental economics and its implications for comparative politics. *Qualitative Methods*, *6*(2), 17-21.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., … Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, *343*, 30-31.

Monogan, J. E. (2015). Research preregistration in political science: The case, counterarguments, and a response to critiques. *Political Science & Politics*, *48*, 425-429.

Moravcsik, A. (2014). Transparency: The revolution in qualitative research. *Political Science in Politics*, *47*, 48-53.

Neumark, D. (2001). The employment effects of minimum wages: Evidence from a prespecified research design the employment effects of minimum wages. *Industrial Relation*, *40*, 121-144.

Nielsen, R., Findley, M. G., Candland, T., Davis, Z., & Nielson, D. L. (2011). Foreign aid shocks as a cause of violent armed conflict. *American Journal of Political Science*, *55*, 219-232.

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*. Retrieved from http://www.nature.com/news/scientific-method-statistical-errors-1.14700

Nyhan, B. (2015). Increasing the credibility of political science research: A proposal for journal reforms. *PS: Political Science and Politics 48*(S1): 78–83.

O'Connor, A. (2016, April 13). A decades-old study, rediscovered, challenges advice on saturated fat. *The New York Times Well*. Retrieved from http://nyti.ms/1SynN1M

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

Pepinsky, T. B. (2015). Context and method in southeast Asian politics. *Pacific Affairs*, *87*, 441-461.

Rayhill, M., Sharon, R., Burch, R., & Loder, E. (2015). Registration status and outcome reporting of trials published in core headache medicine journals, *Neurology*, *85*, 1789-1794. doi:10.1212/WNL.0000000000002127

Reinhart, C. M., & Rogoff, K. S. (2010). Growth in a time of debt. *American Economic Review: Papers and Proceedings*, *100*, 573-578.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, 534-547.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016

Sterling, T. (1959). Publication decisions and their possible effects on inferences drawn from statistical tests—Or vice versa. *Journal of the American Statistical Association*, *54*, 30-34.

Trouble at the lab. (2013, October 19). *The Economist*. Retrieved from http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble

Welzel, C., & Inglehart, R. (2009). Mass beliefs and democratic institutions. In C. Boix & S. C. Stokes (Eds.), *The Oxford handbook of comparative politics* (pp. 297-316). New York, NY: Oxford University Press.

Yom, S. (2015). From methodology to practice: Inductive iteration in comparative research. *Comparative Political Studies*, *48*, 616-644.

Yong, E. (2012). Nobel laureate challenges psychologists to clean up their act. *Nature*. Retrieved from http://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535

Author Biographies

Michael G. Findley is associate professor in the Department of Government at University of Texas at Austin. His research addresses political violence, international development, and international law, and has appeared in *Cambridge University Press, American Journal of Political Science, International Organization*, among others.

Nathan M. Jensen is professor in the Department of Government at University of Texas at Austin. His research interests include multinational enterprises and political risk, the relationship between foreign direct investment and corruption, and tax competition for investment.

Edmund J. Malesky is associate professor in the Department of Political Science at Duke University. His research interests include the political development in Vietnam and China, comparative political economy in Southeast Asia, as well as economic transitions in developing economies. His work appears in the *American Political Science Review, Journal of Politics*, and other venues.

Thomas B. Pepinsky is associate professor in the Department of Government at Cornell University. Currently, he is working on issues relating to Islam, politics, and political economy in Southeast Asia and beyond. His work has appeared in *Cambridge University Press, American Journal of Political Science, World Politics*, and other venues.

Downloaded from cps.sagepub.com at CORNELL UNIV on December 9, 2016