Science Deserves Better:
The Imperative to Share Complete Replication Files

Allan Dafoe

Department of Political Science, Yale University

First Draft: 2012-12-12
This Draft: 2013-12-03

Intended for a 2014 Symposium
in *PS: Political Science & Politics* on *Data Access and Research Transparency*.

Abstract

Scientific knowledge is only as reliable as the empirical analysis on which it is based. For the majority of published statistical analyses, readers have to trust that the scholars correctly implemented the many stages of analysis between primary data collection and the presentation of results—including data cleaning, merging, recoding and transforming, analysis, and output. I advocate the adoption of a simple transparency maxim: *good research involves publishing complete replication files, making every step of research as explicit and reproducible as is practical.* Benefits of replication transparency include: raising the quality and refutability of scientific inferences; dissemination of useful data and code; greater freedom for scientists to explore each others' results and data; and a richer scientific conversation. More transparent replication practices are a scientific public good: the costs are small but borne by the authors; the benefits are great and shared by the broader scientific community and public. Strong norms of replication transparency, once established, are likely to be partly self-enforcing. Data is presented on replication practices and the replicability of published works in political science. I outline good replication practices for scholars, and offer recommendations to journals, universities, and funders.

Comments welcome: allan.dafoe@yale.edu. Replication Files available at: http://hdl.handle.net/1902.1/22160
For input I thank Natalia Bueno, Mats Hammerström, Nicole Janz, Gary King, David Laitin, Daniel Masterson, Gwyneth McClendon, Guadalupe Tuñón, Rich Wilson, Baobao Zhang, Magnus Öberg, and especially the Editorial Committee at the *Journal of Peace Research*, John Bullock, Don Green, and Arthur Lupia.

Contents

1 Introduction

In April 2013 a controversy arose when a working paper (Herndon, Ash and Pollin, 2013) claimed to show serious errors in a highly cited and influential economics paper by Carmen Reinhart and Kenneth Rogoff (2010). The Reinhart and Rogoff paper had come to serve as authoritative evidence in elite conversations (Krugman, 2013) that high levels of debt, especially above the "90 percent [debt/GDP] threshold" (Reinhart and Rogoff, 2010, 577), posed a risk to economic growth. Much of the coverage of this controversy focussed on an error that was a "perfect made-for-TV mistake" (Stevenson and Wolfers, 2013) involving a simple error in the formula used in their Excel calculations. The real story here, however, is that it took three years for this error and other issues to be discovered because replication files were not publicly available, nor were they provided to scholars when asked. If professional norms or the *American Economic Review* had required that authors publish replication files this debate would be advanced by three years and discussions about austerity policies would have been based on a more clear-sighted appraisal of the evidence.

An essential characteristic of science is the commitment to transparency: assumptions should be clearly stated, evidence should be publicly verifiable, the basis for inferences should be explicit. Independent researchers should be able to replicate, at least in principle, the structure of inferences linking assumptions, prior theory, other findings, data collection, data processing, and data analysis, to an alleged scientific finding. Much of the institution of science exists to promote transparency, such as the strong norms around citations, the requirement to describe methods, the esteem for formal methods of inference (statistics and formal theory), expectations about maintaining (laboratory or field) notebooks, the expectation to publish proofs of theorems, and the opprobrium reserved for non-reproducible results.

Transparency is a foundation for a number of core features of science: **refutability**, **openness**, **cumulation**, and **minimal barriers to entry**. **Refutability**: Transparency makes scientific work more refutable—more subject to detailed criticism—which is the basis for scientific progress. **Openness**: Transparency makes a scientific enterprise more open to exploration by others, facilitating divergent interpretations of results and alternative uses of various contributions. **Cumulation**: The cumulation of data, tools, and findings is essential for the progress of science. The sharing of findings is incentivized through rewards to publications. There are not, however, adequate individual incentives for the sharing of data and tools. The sharing of data and tools are an essential public good for science; a commitment to transparency would promote this public good. **Minimal barriers to entry**: Transparency, by making more steps of the scientific process publicly observable, makes it easier for students and novices to learn the art of particular fields of science. Keeping scientific discussion accessible improves the scientific enterprise by reducing the costs to training new scientists, by bringing in new perspectives, by permitting more cross-disciplinary conversation, and by keeping fields open to external criticism.

This paper makes a simple argument. Political science should take its commitment to transparency more seriously by insisting that researchers *publish complete replication files, making every step of research as explicit and reproducible as is practical.* In return, political science will become more refutable, open, cumulative, and accessible. Science deserves this commitment from us.

This paper proceeds as follows. Section 2 reviews some evidence about the current state of replication practices in political science. Section 3 elaborates on the benefits of greater transparency through the sharing complete replication files. Section 4 makes specific recommendations for authors, journals, and universities. Online Appendix A discusses exceptions to the prior recommendations for confidential, costly, or proprietary data or code. Online Appendix B discusses a proposal for Replication Audits.

The recommendations of this paper apply to any domain of science in which some feature of inference could be practically made more explicit and reproducible. These recommendations apply especially to modes of inference that use computers, since any processing involving a computer can be codified and made reproducible. For this reason this paper focusses primarily on replication practices in statistical studies, though the recommendations apply equally to computational theory (theoretical models using computer simulations or solutions). Non-computational modes of inference can also be made more transparent. For example, Moravcsik (2010) offers valuable recommendations to qualitative researchers.

2 Evidence from Political Science

What is the current state of replication practices in political science? Gherghina and Katsanidou (2013) found that only 18 of 120 political science journals have a replication policy *posted* on their websites, to say nothing about enforcing those policies. To provide additional data on the state of replication practices in political science, I collected data on: (1) the availability of replication files for publications in two leading journals—*APSR* and *AJPS*—and (2) scholars' attempts to replicate publications.

2.1 Replication Practices at *APSR* and *AJPS*

APSR does not have a policy of requiring replication files, though it encourages them. For example, in the submission guidelines it states that authors "are expected to address the issue of data availability. You must normally indicate both where (online) you will deposit the information that is necessary to reproduce the numerical results and when that information will be posted." By contrast, the policy at *AJPS* as of 2010 states that articles "will not be published unless the first footnote explicitly states where the data used in the study can be obtained", the acceptance letter provides instructions for posting files to *AJPS*'s Dataverse site, and the editor, Rick Wilson, frequently double checks that files are posted and has held up publications that have not posted replication files. The beneficial effects of this policy and editorial involvement are evident in Figure 1.

Data was collected[1] on the availability of replication files for recent publications in the two top political science journals, the *American Political Science Review* since 2010 and the *American Journal of Political Science* since 2009. We found that 48% of publications employing statistical analysis stated on their first page that replication files were available; we were able to find replication files for 68% of these. We were also able to find replication files for 18% of the publications that did not state that replication files were available.

As Figure 1 shows, publications at *APSR* are much less likely than *AJPS* to state that replication files are available, and somewhat less likely to provide replication files. Since 2011, nearly 100% of publications at *AJPS* state that replication files are available, increasing dramatically from 2009 before the new replication policy. This policy seems to have substantially increased the actual availability of replication files, though we were still unable to find replication files for about 35% of the publications in *AJPS* 2011-2013.

[1] Data was collected by Guadalupe Tuñón, Peter Repucci, and myself.

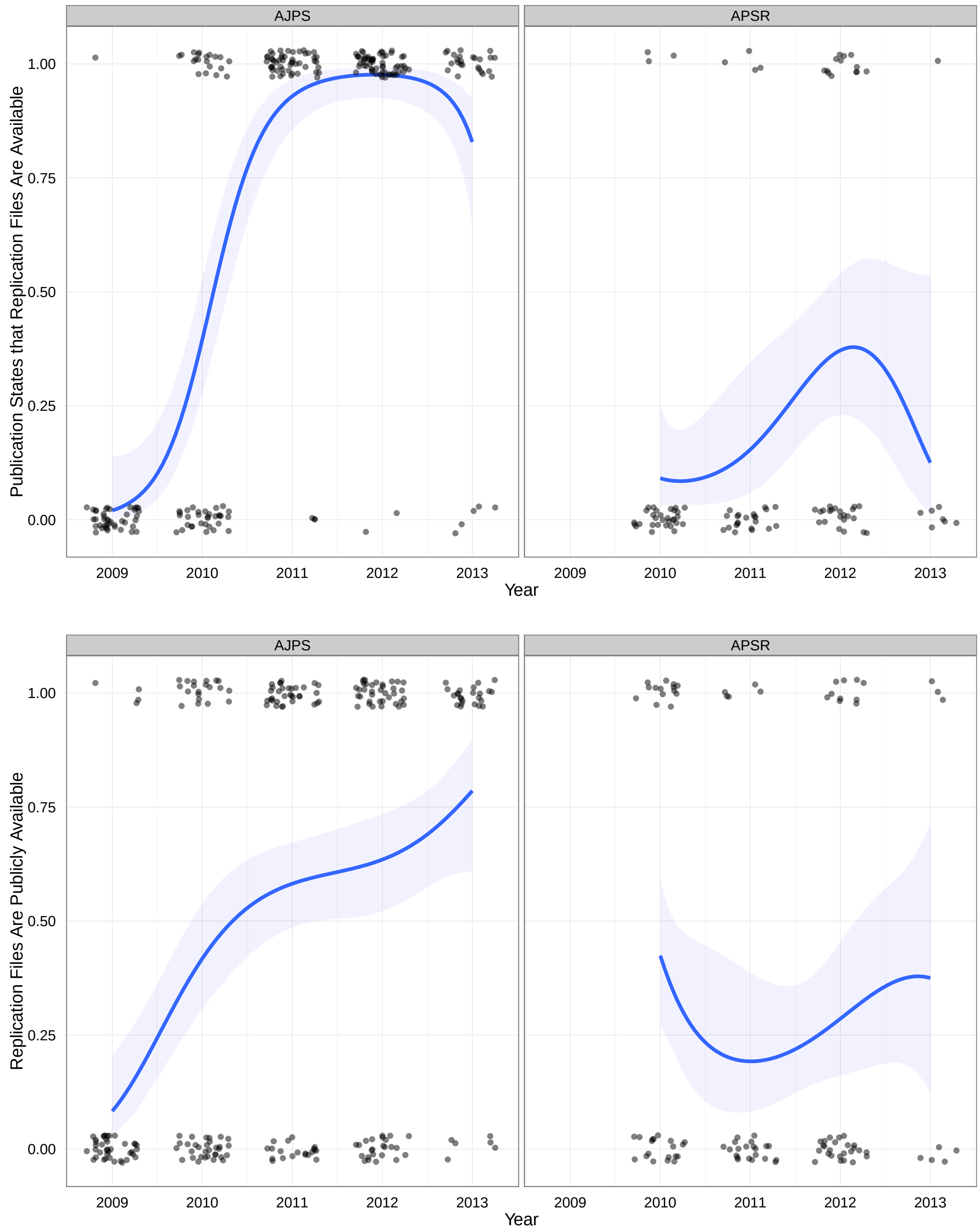

Figure 1: Proportion of published articles employing statistical analysis in *APSR* and *AJPS* that (top row) state that replication files are publicly available, and (bottom row) for which replication files are publicly available. Confidence intervals provide an assessment of whether observed proportions could have come from the same underlying distribution. Results: 30% of the publications at *APSR* provided replication files, 65% of the publications at *AJPS* 2011-2013 provided replication files; *AJPS*'s 2010 policy substantially increased provision of replication files.

2.2 Robustness of Published Results to Replication

The above data does not speak to the quality of the replication files that are provided, nor the actual robustness of results. Addressing this question would require a more systematic evaluation of the literature, such as from a Replication Audit (see Online Appendix B).

To offer some preliminary data on this question I surveyed three groups of scholars about their experiences attempting to replicate statistical studies; these groups were students from my Ph.D. Methods class, students from Gary King's Ph.D. methods class, and subscribers to the Political Methodology listserve. These numbers should be interpreted with caution since it is not from a representative sample: respondents selected into the survey, and respondents selected the work they wanted to replicate. See Online Appendix D for more details about this survey.

This data suggests a mixed conclusion. On the one hand, of those who responded to the reproducibility of the result, about 52% reported that they were "able to precisely reproduce the main results" and only 13% reported that they were "not able to approximately reproduce the main results." This suggests that many results in political science can be, at least superficially, reproduced, but also that many seem only approximately reproducible. Of those who responded to the robustness of the results, 36% reported that "most or all of the key results were robust", 20% that there were "major technical errors though these didn't change the main results", and 56% that results were not robust (responses 5-7). This is encouraging in how many results were found to be robust, while also reinforcing the value of strong transparency norms so that the many fragile results can be more easily uncovered and examined. Also, given that more reproducible and robust work is more likely to share replication files (Piwowar, Day and Fridsma, 2007), these numbers are probably an optimistic appraisal of the reproducibility and robustness of statistical work in political science.

3 Benefits of Sharing Complete Replication Files

This section will discuss some of the specific benefits of the sharing of complete replication files, which are defined as files that make as explicit and objective as practical every step of research from initial data collection to final statistical output.[2] These benefits include greater refutability, openness, cumulation, and ease of overcoming barriers to entry.

A primary benefit of sharing replication files is that it makes research more *refutable*, and therefore makes the body of non-refuted findings more informative. Fragile, misleading, and non-replicable[3] statistical analyses can be largely eliminated by the simple requirement that authors be required to submit complete replication files before publication.[4] Doing so will deter many scholars from publishing unreliable analyses, and the scientific community can be relied upon to expose many of those remaining.

Consider the cases of serious fraud that have been uncovered in psychology. Uri Simonsohn used data analysis techniques (Shea, 2012; Simonsohn, 2012) to detect suspicious data patterns in psychology, and has uncovered incidents of fraudulent data which has led to multiple retractions of

[2] See section 4.1 for more discussion of what constitutes "complete replication files".

[3] We might distinguish between the "replication of a study" in which the research design is replicated on *new* data (also called "broad replication"), and the much less informative but nonetheless important "replication of an analysis" (or "narrow replication") in which the analysis is replicated on the *same* data. Sharing of replication files foremost promotes replications of analysis, though it might also promote replications of studies if the greater transparency facilitates the execution of the study on a new sample.

[4] This paper focusses on replication practices for statistical empirical work. Qualitative scholarship would also be much improved by the adoption of stronger practices of data transparency (see Moravcsik 2010).

articles and the resignation of prominent tenured professors. One psychologist was found guilty of fabricating data for over 50 publications; this professor didn't just "massage" the data, or report only convenient analyses, he literally made the data up and then gave it to his students to analyze for their dissertations (Bhattacharjee 2013). Simonsohn (2012) argues that "requiring authors to post the raw data" will "make fraud much less likely to go undetected".

Fraudulent science probably only makes up a tiny proportion of the scientific output. However, it threatens to dramatically reduce the public credibility of science. Of potentially greater concern (Stevenson and Wolfers, 2013) is the unknown proportion of fragile, misleading, or non-replicable results. The data reported in 2.2 suggests this proportion is not trivially small in political science.

The sharing of complete replication files has the additional benefit that it *opens up* scientific research to the questions, insights, and exploration of others. Rather than confine reviewers and readers to a snapshot of the data that has been carefully prepared by the authors, a reader of an empirical analysis with a question or insight could immediately go to the data to evaluate it. Readers might want to double check that primary data has been merged correctly, evaluate how known coding issues have been addressed, identify influential observations, examine particular subsets of the data, implement an alternative conditioning strategy or estimator, or plot the data in potentially insightful ways. Of course, as with any analysis, any findings from such a reanalysis should follow from a principled and well-motivated empirical strategy, and scholars should be wary of the biases from multiple-comparisons (data dredging).

The full set of questions that a reader might have is vast, and cannot, even in principle, be answered in the text of a paper or supplementary materials. However, authors need not anticipate every possible question. Instead, they need only make a sufficient effort to warrant publication, and provide clear and complete replication files so that the scientific community is able to evaluate and build upon their work. The scientific paper is currently a snapshot of a data landscape. Instead, the scientific paper should become an open safari of the data landscape, from which the reader is encouraged to depart at any time and explore the landscape on her own. Providing that level of freedom would allow scholars to explore and build upon the data to answer their own questions, and otherwise to be much more active participants in the scientific journey of the authors.

Scientific productivity exhibits network externalities. The *cumulation* of data and tools provides substantial benefits beyond the intentions of the creators. A scholar will collect some data or code a variable for some specific research purpose; however, often those same data can be used to answer other questions. This is especially the case for those kinds of observational data, such as cross-national data, that are valued by multiple research programs. For example, the statistical study of international relations has benefited greatly from the creation of large standardized datasets of country and dyad characteristics and behavior that have cumulated from the work of hundreds of scholars. These externalities are also present in experimental research. Experimental manipulations can be "reused" for *down-stream experiments* in which other causal effects are investigated, and novel hypotheses can often be tested by examining causal effects within particular sub-groups.

Sharing complete replication files is also likely to incentivize scholars to be more careful: (e.g., see Andrew Gelman here): we face trade-offs in how we invest our time and we are likely to invest more effort in those stages of our research that are most subject to scrutiny and judgement. By making more of the research process subject to scrutiny, scholars will have greater incentives to be cautious with those parts of the research process. This incentivizing effect may be one of the primary benefits of stronger replication norms.

Sharing code for analysis and presentation *lowers the barrier to entry* for students and others, and promotes the dissemination of useful techniques. Students especially benefit from having access to replication code since it allows them to see precisely how prominent scholars execute their empirical analyses and provides an opportunity for junior scholars to contribute to the research frontier

(Rich, 2013). Instead of needing to “study under" or collaborate with leading scholars to learn their statistical methods, scholars will be able to learn by working through replication files.

Papers that share their replication data and code have greater visibility and more citations (Gleditsch, Metelits and vard Strand, 2003; Piwowar, Day and Fridsma, 2007). This is probably partly a selection effect, but also probably partly a causal effect. It is much harder to build off of a study for which replication files are not available. In addition, the sharing of replication files provides a public signal about the quality, confidence, and professionalism of a scholar.

Transparent replication practices are a scientific public good: the benefits are large and shared by many, the costs are small but born largely by the authors. While the benefits vastly outweigh the costs, transparency will likely be under-provided unless individual's incentives are aligned with the group's. Strengthening of formal incentives could help, such as if journals, funders, and universities insist that replication files be publicly posted as part of the publication, funding, and promotion process. Ultimately, however, scientific practice follows scientific norms. To incentivize adequate transparency we need to broadly promote transparency norms such as is articulated in the most recent Guide to Professional Ethics in Political Science : “openness is an indispensable element of credible research and rigorous analysis, and hence essential to both making and demonstrating scientific progress." Specifically, I recommend the following transparency maxim for computational work:

> Transparency Maxim: *Good research involves publishing complete replication files, making every step of research as explicit and reproducible as is practical.*

The transparency maxim is both a descriptive statement that good research tends to publish replication files, and a normative statement that the publishing of replication files is a necessary component of good research.

The transparency maxim is likely to be partly self-enforcing. Researchers are more likely to publish complete replication files as they are more technically proficient, more concerned about the quality of their work, more confident in their work, more concerned with the scientific enterprise, and more concerned with being perceived as producing good research. These motives generate a correlation between replication files and good research. For example, Wicherts, Bakker and Molenaar (2011) report that willingness to share data is positively associated with the strength of the evidence and the quality of the reporting of statistical results. Similarly, some scholars adopt judgmental heuristics based on the availability of replication files (e.g. here).

As this descriptive association becomes stronger, publishing replication files will send a positive signal about the quality of one's research (or the failure to publish replication files will send a negative signal). Low quality research cannot easily “fake" this signal since the very act of publishing replication files makes it much easier to evaluate the quality of the research. This signal will then encourage scholars, journals, and universities that produce good research and wish to be perceived as such to publish replication files. Publishing replication files is thus an informative signal of the quality of one's research.

4 Recommendations

The following are some specific recommendations about how to produce good replication files for researchers engaged in statistical analysis. This advice is similarly applicable to scholars engaged in computational theory. Some advice is also offered here for journals, universities, and funders about how best to promote these practices. The *American Political Science Association* has also recently revised its *Guide to Professional Ethics, Rights and Freedoms* to emphasize and clarify researchers' "ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated" and offers additional guidance on these topics for quantitative and qualitative researchers. For advice for qualitative researchers, see Moravcsik (2010). Davenport and Moore (2013), Hook et al. (2010), and ICPSR offer advice on data preparation and archiving.

4.1 Recommendations for Statistical Studies

1. **Do all data preparation and analysis in code.** Even if analysis is done by "clicking and pointing", most statistics programs (such as Stata and SPSS) produce the code required to replicate each step.

2. **Adopt best practice for coding.** Some recommendations to keep in mind are:

   - **Use comments and functions to make your code clear.** Keep your code clean and clear. Comment liberally to remind yourself and communicate to others what your code is (supposed to be) doing. Use functions to execute specific commands, especially when these commands are repeatedly used.
   - **Test your code.** Build in routine tests to make sure that your code is doing what you think it is doing. Execute the same procedure in multiple redundant ways to reduce the risks of a mistake. It is not uncommon for results to be driven by a misimplemented routine such as the mishandling of missing data.
   - **Run your final code all the way through from scratch.** Before finalizing the paper, the entire replication code should be rerun from beginning to end. Make sure you set a seed, and make a *log* file (in Stata) or use something like *knitr* (in R), so that this final run is recorded and fully reproducible. Also make sure that all relevant data files are included in the replication files. One way to ensure this is to only call data files from within the folder where the code is stored, and then to upload the entire replication folder for archiving.
   - For a helpful discussion of strategies to improve replication practices and code, see (Bowers, 2011) and Appendix A of (Shalizi, 2013).

3. **Build all analysis from primary data files.** Download data files from the original source, and include a precise reference in your code or paper to this original source. Lock these primary files to prevent accidental changes to them. This way any errors that occur will take place in your code, which can be diagnosed and corrected. Share these original data files along with all other files in your replication files. What good is sharing a final data file and replication code, as many scholars currently do, if the crucial decisions and errors were made earlier in the merging and cleaning of the data? We should take the attitude towards a statistical estimate as legal courts take towards evidence: there should be a clearly documented "chain of custody" from trusted primary files to the final reported output. In our case, this chain of custody should be fully documented and easily reproducible.

4. **Fully describe your variables.** Somewhere—in variable labels, a codebook, paper, or comments in the code—the meaning of variables needs to be clearly communicated. The original sources or coding rules for variables should be provided. A reader should be able to trace a variable back to its original creation, and the author who first created a variable

should clearly document the rules by which the variable was constructed. It is unacceptable to share data files for which it is unclear precisely where a variable came from, let alone one in which the variables names are an indecipherable character string.

5. **Document every empirical claim.** Every empirical claim in a paper based on the data should be explicitly produced somewhere in the replication code. This includes all graphics and tables, but also any in-text reference to some feature of the data. It should be easy for others to link empirical claims in the paper to the relevant portion of the code. One way to do this is to include a quote in the replication code of the sentence in the paper in which the empirical claim is made; this way a reader can search for the text of the empirical claim in the replication code. Another strategy is to have one's replication code follow the structure of the paper.

6. **Archive your files.** Upload this finalized set of analysis code and data files to a reliable third-party site such as ICPSR or Dataverse.

7. **Encourage coauthors to adopt these standards.** Maintaining good replication practices is more difficult when one's collaborators do not have the same replication practices. Scholars may be understandably reluctant to impose strict requirements on their colleagues. I recommend sharing with them papers such as this one that outline good replication practices and the reasons behind them, and otherwise leading by example.

4.2 Recommendations for Journals

Journals are the key site for improving replication practices. I describe below a set of policies that journals could readily adopt. This package of policies was crafted to have minimal cost to the journal, to maintain the status quo with respect to when data and code should be shared with the public (at time of publication), and to maximally improve replication practices. Adopting these recommendations will improve the quality of work being published, will signal that the journal has higher standards, and will likely increase the prominence of the journal (Gherghina and Katsanidou, 2013).

1. **Require complete replication files before acceptance.** The simplest policy is a nominal requirement that the authors make replication files available. For example, *AJPS* requires that "the first footnote explicitly states where the data used in the study can be obtained for purposes of replication." As is suggested by the data in Figure 1, this change made a substantial difference for *AJPS* (compare years before and after 2010). However, despite that nearly 100% of *AJPS* articles now explicitly state the location for replication files, less than 75% actually provide replication files. Others have similarly noted the limits of requiring authors to agree to or sign statements of intent to share data (Wicherts et al., 2006; Savage and Vickers, 2009). As such, to achieve sufficiently high compliance, journals need to actually ensure that replication files are posted or to post it themselves. For example, *the Quarterly Journal of Political Science* ensures that replication files are available and that it is possible to replicate the results before an article is published. This process of checking or posting replication files can also be automated in the workflow programs used by journals. Journals may want to adopt the policy at *Nature* of requiring an "accession number" or URL for the replication files at the time of first submission; the files can then be released at the time of acceptance.

2. **Encourage high standards for replication files.** The journal should articulate its expectations about the quality of replication files to authors. Ideally the journal will encourage high standards, such as those articulated in section 4.1. Replication files could be made available to reviewers after an R&R, allowing reviewers the option to include the quality of replication files in their assessment of the publication.[5]

3. **Implement a Replication Audit.** A Replication Audit involves assembling a Replication Team of trusted researchers to evaluate the replicability and robustness of a random subset of publications from the journal. By guaranteeing regular space in the journal for the Replication Audit the journal (1) helps reward the act of evaluating the replicability and robustness of published work, (2) incentivizes authors to invest additional effort to make sure that their results are replicable, robust, and that their inferences are not misleading, and (3) provides a diagnostic of the (hopefully improving) quality of empirical work in the journal. The Replication Audit is described in Online Appendix B. A Replication Audit is preferred to the exclusive publishing of replication articles on a case-by-case basis because the latter process is more susceptible to publication bias that will over-represent “interesting" replications that claim to overturn earlier studies.

4. **Retract publications with non-replicable analyses.**[6] If an analysis can not be replicated, even by the original authors when given ample opportunity to do so, the results from the study cannot be trusted and the study should no longer be a part of the public scientific record. Publications based on non-replicable analyses should, therefore, be retracted. David Laitin, in a personal communication, recommended the adoption of a standard retraction procedure for political science. Doing so would make the retraction policy more transparent, remove editorial discretion, and insulate editors from legal retaliation. The standards for such a retraction policy could be set so that only the most egregious cases of non-replicable analyses are retracted: where the finding cannot be approximately reproduced by the original authors. Even if rarely activated, a retraction policy would promote replication practices by establishing the norm that authors are responsible for providing adequate replication files and by providing strong sanctions against the worst kinds of non-replicability.

4.3 Recommendations for Universities

1. Universities can provide institutional support for producing complete replication files. For example, the Institution for Social and Policy Studies (ISPS) at Yale University provides a service in which they help produce and publish complete replication files in both R and Stata for ISPS funded research (see here). Harvard’s Institute for Quantitative Social Science has built the Dataverse Network Project which provides long-term archiving of replication files and other services.

2. Universities could encourage and expect high replication standards from their students and scholars. Students should be encouraged to submit replication files for course papers. Departments could have a policy recommending publication of complete replication files for all published work.

[5]Another option is to require that reviewers have access to replication files from first submission. However, a number of scholars have expressed concern over this proposal because they are not comfortable with others having access to replication materials so much in advance of publication.

[6]I thank David Laitin for raising this idea.

3. Norms of scholarly evaluation could place more emphasis on transparency, and specifically the provision of replication files.

4.4 Recommendations for Funders

1. Require recipients of funding to commit to transparency, and specifically to publish replication files. The *NSF* for example now requires a data management plan as part of any proposal, though of course this is not sufficient (for a satire of one, see here).

5 Conclusion

The study of politics rightly aspires to be scientific: it aims to establish generalizable causal insights from the non-subjective, replicable, and transparent empirical evaluation of precise and logical theories (Gerring, 2011, 11). Relative to many natural sciences, however, political science faces daunting methodological challenges. We are unable to use experiments to resolve many of our questions. We can rarely isolate mechanisms and processes in a controlled setting. The behavior of our subjects rarely follow simple mathematical patterns, are highly context dependent, and adapts to our interventions and theories. However, political science has better replication practices than many of the natural and social sciences. We should be proud of this and continue to lead the way.[7]

[7]One site for continuing this conversation is at a blog precisely on this topic, set up by Nicole Janz: http://politicalsciencereplication.wordpress.com/

References

Bowers, Jake. 2011. "Six Steps to a Better Relationship with your Future Self." *Political Methodologist* 18(2).

Davenport, Christian and Will H. Moore. 2013. "Conflict Consortium Data Standards & Practices.".

Gerring, John. 2011. *Social Science Methodology: A Unified Framework*. Cambridge University Press.

Gherghina, Sergiu and Alexia Katsanidou. 2013. "Data Availability in Political Science Journals." *European Political Science* .

Gleditsch, Nils Petter, Claire Metelits and Håvard Strand. 2003. "Posting Your Data: Will You Be Scooped or Will You Be Famous?" *International Studies Perspectives* 4(1):89–95.
URL: *gking.harvard.edu/files/replvdc.pdf*

Herndon, Thomas, Michael Ash and Robert Pollin. 2013. "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff." *Political Economy Research Institute Working Paper Series* (322).

Hook, Les A., Suresh K. Santhana Vanna, Tammy W. Beaty, Robert B. Cook and Bruce E. Wilson. 2010. "Best Practices for Preparing Environmental Data Sets to Share and Archive.".
URL: *daac.ornl.gov/PI/BestPractices-2010.pdf*

Krugman, Paul. 2013. "The Excel Depression." *The New York Times* .
URL: *http://www.nytimes.com/2013/04/19/opinion/krugman-the-excel-depression.html*

Moravcsik, Andrew. 2010. "Active Citation: A Precondition for Replicable Qualitative Research." *PS: Political Science and Politics* 43(01):29–35.

Piwowar, Heather A, Roger S Day and Douglas B Fridsma. 2007. "Sharing Detailed Research Data Is Associated with Increased Citation Rate." *PLOSOne* 2(3):e308.

Reinhart, Carmen M. and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review: Papers & Proceedings* 100(2):573–78.

Rich, Timothy S. 2013. "Publishing as a Graduate Student: A Quick and (Hopefully) Painless Guide to Establishing Yourself as a Scholar." *PS: Political Science & Politics* 46(2):376–379.

Savage, C J and A J Vickers. 2009. "Empirical study of data sharing by authors publishing in PLoS journals." *PLOSOne* 4(9):e7078.

Shalizi, Cosma Rohilla. 2013. *Advanced Data Analysis from an Elementary Point of View*.
URL: *http://www.stat.cmu.edu/∼cshalizi/ADAfaEPoV/*

Shea, Christopher. 2012. "The Data Vigilante." *The Atlantic* December.
URL: *http://www.theatlantic.com/magazine/archive/2012/12/the-data-vigilante/309172/*

Simonsohn, Uri. 2012. "Just post it: The lesson from two cases of fabricated data detected by statistics alone." pp. 1–30.

Stevenson, Betsey and Justin Wolfers. 2013. "Six Ways to Separate Lies from Statistics." *Bloomberg* .
URL: *http://www.bloomberg.com/news/2013-05-01/six-ways-to-separate-lies-from-statistics.html*

Wicherts, Jelte M., Denny Borsboom, Judith Kats and Dylan Molenaar. 2006. "The poor availability of psychological research data for reanalysis." *American Psychologist* 61(7):726–728.

Wicherts, Jelte M., Marjan Bakker and Dylan Molenaar. 2011. "Willingness to share research data is related to the strength of the evidence and the quality of statistical results." *PLOSOne* pp. 1–24.

A Exceptions for Data or Code that is Confidential, Proprietary, or a Costly Original Contribution

The recommendations for journals in 4.2 applies to cases where the data and code are not proprietary or confidential. If data or code is confidential, proprietary, or represents a very costly original contribution, the appropriate replication norms need to be more nuanced. As long as scholars and journal editors make a sincere effort to work through these issues we will be in a much improved situation from where we are now. However, I outline below some specific suggestions.

A.1 Confidentiality and Proprietary Data

Some data has issues of confidentiality. Protection of human subjects may require that individual respondents cannot be identified from the data. The authors may have agreed to certain restrictions in order to access their data.

There are a range of potential responses, from least to most limiting in transparency. Authors could strive to be as transparent as possible, and explain to journals and readers the limits on transparency that they choose.

Potential responses include, from less to more restrictive:

1. Remove or scramble identifying labels. Often ID labels are unnecessary for analysis, and after removal individuals can no longer be identified.
2. Remove or scramble other identifying variables, such as I.P. address, age, etc...
3. Authors could share the code that they used to carry out the analysis and a dummy dataset. A dummy dataset is a dataset that involves fake data, but otherwise tries to preserve some features of the data. For example, a version of the actual dataset could be provided in which each variable that may help identify individuals is randomly resorted; this would allow readers to investigate questions about the distributions of the confidential variables.
4. Share other non-identifying statistics in the online appendix, such as density plots and covariance matrices.

For proprietary data, the authors could still provide complete replication code and explain how one can acquire access to the data. Often creative solutions exist for restricted data; some military data is not allowed to be shared with the public, but may be shared more restrictively through military libraries. And as with confidential information, as much information could be provided as possible, such as the covariance matrix and a dummy version of the dataset.

A.2 Original Costly Data (or Code) Acquired by the Authors

Our profession does not reward the collection and provision of data adequately. Most scholars who put together new data understandably want to have restricted access to the data for their publication purposes. The issue is similar to that of patent law: as a discipline we want to incentivize scholars to collect valuable new data and our primary instrument for doing so is the opportunity of data collectors to be the first to analyze their data; however, we also want to allow other scholars to access the data to evaluate the claims of the authors and to contribute to other projects. Currently APSA's Data Access and Research Transparency guidelines (section 6.6) recommends a one year embargo period from the time of publication. Note that even with an embargo, journals could still require that complete replication files be provided to the journal before acceptance. Similar embargo possibilities may be permitted for original costly contributions in code.

B Replication Audit

The idea of a "Replication Audit" emerged from conversations at the Editorial Committee Meeting at the *Journal of Peace Research*. The group agreed that strengthening replication practices was an important goal. One way to do this would be for journals to guarantee space for replications, since this would (1) reward the practice of engaging in replication, (2) provide a sense for the reliability of results, and (3) incentivize authors to be more careful in their analyses.

However, this kind of exercise is susceptible to publication bias: if editors prefer to publish replications that are "interesting", then there will be a bias against replications that find support for the original results, and towards those that claim to overturn the original results. This risks promoting a-theoretic fiddling with analyses in a search for "interesting" results and exaggerated interpretation of results. To address this concern we propose a more systematic approach to replication: the Replication Audit. We recommend that journals commit to a periodic Replication Audit of their publications.

The Replication Audit involves randomly selecting a subset of articles from a set of journals, examining them in a systematic and moderate manner, and reporting the findings. The Replication Audit should be guaranteed space in journals to reduce incentives to exaggerate findings. The Replication Audit will consist of a team of researchers, and each article will be replicated by two separate scholars who are not aware of each other. The lack of individual credit for particular findings will reduce the incentives for individual scholars to exaggerate their replication findings, and the redundant replication will increase the accountability and reliability of each replication.

A second concern facing scholars wishing to engage in replication is that the original authors might resent the scholars who examined their work. Our proposal mitigates this problem because (1) the articles selected for replication will be selected randomly, blind to the identity of the authors, and (2) the respective contributions of members of the team will be kept private. Only the combined output of the entire team will be published.

B.1 Replication Team

The Replication Team will consist of 3-8 prominent researchers ("advisors") who have access to competent graduate students. Each advisor will lead a team of around 3-8 strong graduate students to perform the replications. One or multiple advisors will lead the replication team (the "coordinators").

B.2 Process

A list of relevant articles will be compiled. For now the relevant scope would be all articles employing statistical analysis or computational theory, within the relevant set of journals during a recent span of time (say 5 years). These articles will be assigned a random number to determine their "priority". The coordinators will go through the articles for each journal, in descending priority, evaluating whether replication files are available. If they are not, an email will be sent to the authors to ask for their replication files, with one follow up email after three weeks. The team will proceed down the list until they have collected N articles with replication files for each journal. Each article will then be assigned to two members of the replication team, so there will be in total $2N$ replication attempts.

B.3 Minimal Replication Standards

The Replication Team will answer the following questions for each article.

- Were replication files publicly available? If not, were replication files available upon request?
  - 0 if not available.
  - 1 if available upon request.
  - 2 if publicly available.
- What was the quality of replication files?
  - 0: A dataset (with all variables needed for the analysis) was available. No analysis code was available.
  - 1: A dataset (with all variables needed for the analysis) and some analysis code were available.
  - 2: A dataset (with all variables needed for the analysis) and complete analysis code were available.
  - 3: Most primary datasets, code to create the final dataset, and complete analysis code were available.
- Nominal Replicability?
  - -1: We were not able to replicate the study due to our technical/software limitations, or overly complex code.[8]
  - 0: We were not able to approximately reproduce the main results.
  - 1: We were able to approximately, but not precisely, reproduce the main results.
  - 2: We were able to precisely reproduce the main results.

B.4 Minimal Robustness Standards

The Replication Team could also evaluate the articles for major technical errors and robustness to arbitrary aspects of the specification.

- Technical Errors?
  - 0: We identified major[9] technical errors.
  - 1: We did not identify any major technical errors.
- Narrow Robustness? For this question, scholars will examine whether the main results are robust to sensible modifications of arbitrary aspects of the specification and any errors in implementation. We will try to layout a systematic approach to this question, though given the complexity of analysis this may not be possible.

[8] If in doubt, the analyst should spend at least 10 hours trying to replicate the study before coding -1.
[9] A major technical error is an aspect of the analysis that could have had or did have potentially serious consequences, and that most experts would agree should not have been done.

- 0: One or more key results were driven by a technical error.
- 1: One or more key results were not robust to a sensible modification of an arbitrary aspect of the specification.
- 2: Most or all of the key results seemed to be robust to technical errors and sensible modifications of arbitrary aspects of the specification.

The analysis done for this Replication Audit will be completely documented in code. We will try to standardize appropriate parts of the replication activities. Members are encouraged to document their reasons for their coding decisions. When two members of the team have divergent codings, we will have them discuss by email their reasoning. Copies of these conversations will be preserved in our replication files, so that our process is as transparent as possible. Advisors will help adjudicate divergent codings.

C Availability of Replication Files

The data about the availability of replication files at *APSR* and *AJPS* was collected as follows by Guadalupe Tuñón (Peter Repucci coded AJPS 2009).

- All publications performing quantitative analyses from *AJPS* and *APSR* since 2010 were downloaded and coded.
- **StateAv**: Does the paper indicate that replication files should be available? Yes or no answer. Yes was given to those papers that indicate data is available upon request or provide an explicit reference to where data can be found.

The subset of papers that did not indicate that replication files should be available were coded as "no" for all subsequent questions. The remainder were coded as follows:

- **UpReq**: Does the paper indicate that data is available upon request? Yes or no answer
- **RepRef**: Does the paper provide a reference to a public source (author's webpage, dataverse, etc) where data can be found? If it did, the reference was included.
- **RepAv**: For those that indicate replication files are publicly available, were they in fact available? Yes if we were able to access the data where the authors claimed it was or after a search in the authors' webpages. (Publications for which data was indicated to be available upon request take a value of "no", since data is not *publicly* available).
- **RepFilesAv**: Are replication files available? An undergraduate RA, Peter Repucci, re-examined all publications for which **StateAv**=No. He searched for replication files on Dataverse, on the authors' websites, and then elsewhere on the internet. If he found replication files he coded **RepFilesAv**=1, if he could not find replication files after 5 minutes he coded **RepFilesAv**=0. Then, in the R code, **RepFilesAv** is set to **RepAv** for all publications in which **StateAv**=Yes.

D Survey of Replication Exercises

This survey is evaluating the extent to which quantitative studies in political science are replicable. Please answer the following questions with respect to a study that you invested at least a few hours in a replication effort. (If you have attempted to replicate more than one, please reload this page when your are done to fill out another survey for each replication effort.)

Thank you for your input,
Allan Dafoe
Yale University

Replication Materials

- ☐ No replication materials were provided.
- ☐ A limited dataset was provided to perform the analysis, but no analysis code.
- ☐ A limited dataset was provided to perform the analysis and some analysis code.
- ☐ A limited dataset was provided to perform the analysis and complete analysis code.
- ☐ Many primary datasets were provided, as well as code to create the final dataset from the primary datasets and code to analyze the data.
- ☐ Other: please explain

>>

Please describe the results of your replication efforts (select all that apply)

- [ ] I was not able to approximately reproduce the main results.
- [ ] I was able to approximately reproduce the main results.
- [ ] I was able to precisely reproduce the main results.
- [ ] I found one or more major technical errors, though these didn't change the main results.
- [ ] I found that one or more key results were driven by a technical error.
- [ ] I found that one or more key results were driven by an arbitrary aspect of their analysis.
- [ ] I found that one or more key results were fragile in a manner that would lead an impartial scholar to substantially discount the value of the original study.
- [ ] Most or all of the key results were robust.

Please describe any other features of your replication exercise that you think might be relevant:

Journal:

Year:

First Author (LastName, FirstName):

Other necessary information to identify the publication:

Respondents were first given the option to select from six characterizations of the availability of replication files. The responses for the full sample are displayed in Figure 2; responses from the PolMeth listserve are similar, but slightly more negative, than those from my class and Gary King's class. Ignoring the "Other" category, 36% of these respondents reported that complete data files and replication code were available (responses 4 or 5), whereas 64% reported that some important element of the replication files were missing (responses 1, 2 or 3). This corresponds closely to the previous finding that 38% of *APSR* and *AJPS* statistical publications make replication files available.

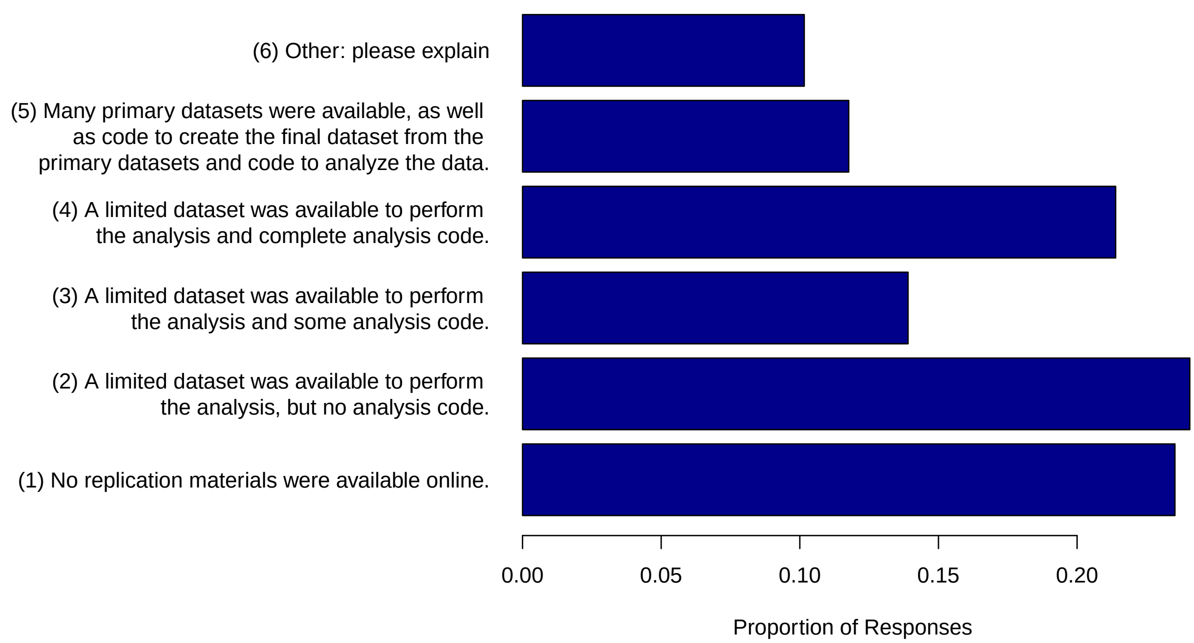

Figure 2: Availability of replication files for respondents.

Respondents were asked to "Please describe the results of your replication efforts (select all that apply)." The responses are displayed in figure 3.

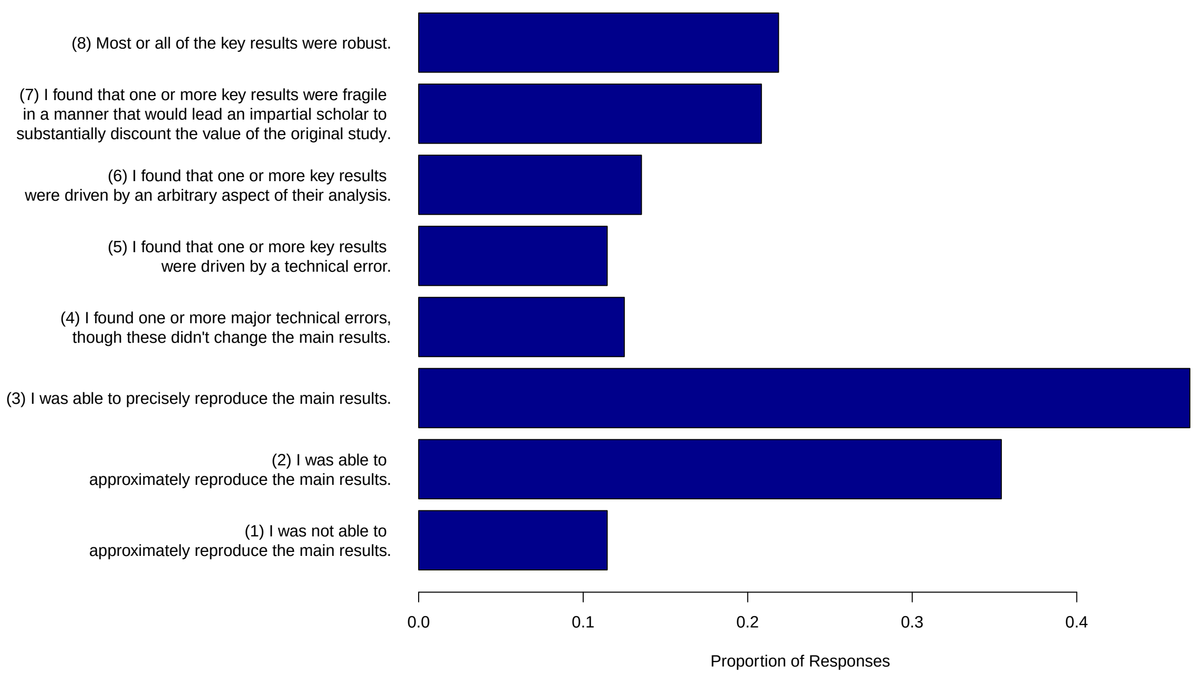

Figure 3: “Please describe the results of your replication efforts (select all that apply)."