

Replication: Why, Where, and How? A Synopsis¹

NILS B. WEIDMANN

University of Konstanz

What is the main motivation for replication in international relations? How can we implement it, and who is in charge? With a variety of proposals put forward by the contributors to this special issue, it seems useful to provide a synopsis. This article does so by asking a set of key questions related to replication, and by summarizing the stance each proposal takes with respect to these questions. This comparison reveals that there are fewer disagreements than we would think: there is broad consensus when it comes to the motivation and the need for replication in our field. However, proposals diverge as to how replication should be best implemented and, to a lesser extent, who should be the driving force behind this. This article concludes with a recommendation for a dual strategy. First, simple replication on the journal side should be implemented as a routine quality check to weed out work that does not even satisfy basic replication standards. Second, journals should be encouraged to allow publication of more comprehensive, in-depth replication studies with an independent contribution.

Keywords: replication, research transparency, peer review

The question of replication in political science and international relations research has been a recurring issue in recent decades. As the articles in the special issue and other recent work illustrate, it is an issue that is far from settled. As such, a continued discussion about appropriate scientific practice is welcome and necessary, since it indicates the willingness of the community to engage in self-reflection and improvement of scientific practice. At the same time, repeated emergence of replication as a topic of discussion could also indicate that this very adaptation has not yet occurred, or if it has, then it has been at too slow a pace.

So where do we stand when it comes to replication? Are there signs of improvement, or is there reason to be pessimistic about the community's ability to self-correct potential shortcomings in the scientific process? How can we adapt replication procedures to recent developments in scientific procedures? The articles in this special issue do a great job of shedding light on these questions. Most importantly, the very existence of these contributions and the ideas presented in them is evidence that critical self-reflection continues to be a part of scientific discourse.

This concluding article of the forum aims to provide a synopsis of the contributions presented. It does so by asking a set of key questions on the motivation and implementation of replication in the field of international relations. The purpose of this overview is not only to relate the ideas and suggestions to each other but

¹The author gratefully acknowledges support from the Alexander von Humboldt Foundation (Sofja Kovalevskaja Award).

also to identify questions and issues where they disagree. This exercise reveals that fewer disagreements exist than we would think; all authors share the motivation to improve the scientific process. However, proposals about the actual implementation of replication differ slightly, and it is useful to take a closer look at the strengths and weaknesses of each of them.

Key Questions about Replication

A synopsis of the work presented in this forum is best done by asking a set of key questions about why and how to implement replication in international relations research. We start by going through the motivations for replication before proceeding to concrete proposals for implementation.

Why Should We Replicate?

A distinguishing feature of scientific analysis is that the “procedures are public” (King, Keohane, and Verba 1994, 8). This means that data, assumptions, analytical procedures, and tools are sufficiently well described such that it is possible for others to understand how the conclusions of the original study were reached. In other words, the original study should be replicable in the sense that it makes all information about the scientific process publicly available. Only then is it possible for others to understand and scrutinize the scientific process and to assess whether or not the original conclusions are valid.

Therefore, the transparency of the scientific process is probably the most important motivation behind replication, and it is a common theme across most articles in this forum. There is a broad consensus among the contributors that transparency is a requirement for scientific findings to be debated and challenged and, thus, a necessary goal to achieve. Thus, the articles argue that replication should be done for the greater good of transparent and sound scientific results. At the same time, attempts to enable this can also be of great value to individual researchers. Results presented in Gleditsch, Metelits, and Strand (2003) confirm that quantitative studies that make their replication data readily available attract more citations. Obviously, lowering the bar for replication can be in researchers’ self-interest.

Why Do We Need to Enforce Transparency?

If most researchers were to believe in the citation-generating effect of replication, we probably would not need entire special issues to be devoted to the topic. However, most authors in this forum would probably not attribute a strong effect to Gleditsch, Metelits, and Strand’s (2003) findings and rather assume that the individual incentives for increasing transparency of one’s own research are low. Colaresi (2016) nicely models the publication process as a game in which the author has private information about the robustness of an empirical result but has incentives not to share this information in order to obtain benefits from publication. This perception is also conveyed in Janz (2016), who proposes to employ replication more widely in graduate courses in order to weed out bad work. Thus, there seems to be some consensus among the contributors that science is not self-healing and that there are incentives for scientists to publish potentially weak findings. For that reason—and there also seems to be some agreement on this—there is a need to externally impose transparency standards. If individual incentives for transparency are too low, mechanisms need to be put in place to enforce compliance. As we will see in what follows, proposals differ when it comes to the actual implementation of this enforcement, but a consensus exists that it needs to be done.

What Is Replication Anyway?

Before moving on to different ways to implement replication, we should take a closer look at what we mean by it. As Martel García (2016) points out, there is still considerable disagreement among researchers. He argues that it is difficult to draw the line between an original study and a replication since most studies build on previous ones and, thus, replicate and extend their results. For the purpose of a brief discussion, it seems useful to depart from King's (2003) distinction that introduces three levels of replication. The first level is the simple reproduction of original results (duplication), something that Martel García (2016) calls "pure replication." The second level extends the first level to assess robustness to alternative specifications, something that comes close to what Martel García (2016) calls "robustness replication." Finally, the third level checks robustness against alternative data, which can include different samples from the same population ("statistical replication," according to Martel García 2016) or from a different population ("external replication," Martel García 2016).

At least when it comes to quantitative research, most proposals for replication focus on levels 1 and 2. Colaresi (2016) presents "prepublication"—the replication of all analyses conditionally accepted for publication—as a way to ensure at least a basic correspondence of empirical results with those later published in an article. Janz (2016) goes a bit further and encourages replication for training purposes, where students should also test different model specifications and, potentially, data sources. More recently, increasing attention has been paid to King's (2003) level 3, data. Dafoe (2014) discusses this question explicitly and argues that it should be possible to replicate not only the statistical results but also the entire data-generation process. Among the articles in this forum, this point is echoed by Katsanidou, Horton, and Jensen (2016), who propose ways to improve transparency in the data that an analysis is based on.

What Replication Material Do We Need?

Closely related to the question of what we should replicate is the question of which material we need for doing so. For quantitative work, King's (2003) list is helpful. Replication material consists of (i) the final data set required for the analysis and (ii) the code to produce tables and figures. In most cases, this will enable level 1 (prepublication) and level 2 replication studies, where alternative model specifications are tested. However, the latter can be made difficult by the fact that alternative specifications require variables other than those included in the (limited) replication data set.

This is why more recent work has pushed for complete replication files to be shared for quantitative work (Dafoe 2014). *Complete* in this case means that the replication code documents not only the final stage of the analysis (i.e., the statistical tables and figures in the article) but the entire data-generation process as well, starting with the original source data sets. This would enable replicators to understand how the data set was constructed, identify potential problems, and include a number of additional variables from these source data sets that were not provided as parts of the replication data set. Similar attempts to improve data transparency include Katsanidou, Horton, and Jensen's (2016) suggestions for improving data presentation and documentation. This approach implies that more detailed information about the data-generation process is shared as part of the replication material.

Although the push for more data transparency and availability certainly addresses an important problem, that problem may not be easy to resolve. First, there is the question about proprietary data that cannot be shared. For example, the study by Pierskalla and Hollenbach (2013) published in the *American Political*

Science Review analyzes the effect of cell phones on violence. The cell phone coverage data, however, are subject to license restrictions, which is why the replication data set does not include the independent variable and is of limited use. Many other data providers may impose similar restrictions, which makes it impossible for authors to share source data. Second, if source data need to be shared along with the first published article using the data, incentives for authors to perform expensive coding will decrease, limiting progress in data collection. Third, data-generation routines often require specialized skills and software as, for example, in spatial analysis (Gleditsch and Weidmann 2012).

Because of the problems that might arise in extending replication to data generation and management, most contributors agree that for quantitative analyses, the replication material should include the replication data set and code to generate tables and figures. For qualitative analysis, replication is considerably more difficult. Some of the articles in this forum provide helpful suggestions for improving transparency at various stages of the research process. Tucker (2016) proposes to apply grounded theory used for theory building in sociology to international relations research. Although this does not imply that more, and other types of, replication material would be distributed along with the article, it does mean that the article itself will be much more detailed when it comes to the statements from which theory is derived.

Who Replicates?

Equipped with a number of suggestions for increasing the transparency of both qualitative and quantitative research, we need to ask who is actually in charge of replication. As we saw in the preceding discussion, the enforcement of transparency through replication means that replicators must be designated since replication efforts are unlikely to emerge spontaneously from within the community. For quantitative studies, where replication presents less of a challenge, different implementations have been proposed. First, Colaresi (2016) proposes to hire graduate students to “preuplicate” every study conditionally accepted for publication in a journal. In fact, this is a procedure already employed by journals, such as *Political Science Research and Methods* and now also by *International Interactions*. Second, Janz’s (2016) approach is to use replication in graduate teaching, where students can freely select studies to be replicated as part of their methods training. A third proposal comes from Dafoe (2014), who suggests establishing a replication audit board at a journal that randomly selects articles to be replicated by graduate students.

In sum, the presented solutions suggest a relatively narrow replication of research articles to be carried out by graduate students. The learning benefits for students are certainly substantial, but there are a number of potential issues to consider, many of which are discussed by Janz (2016). Most of these issues can be circumvented if students are relieved of the task of selecting studies for replication themselves, as is done in Colaresi’s “preuplicate” approach and Dafoe’s audit procedure. Ideally, one could even kill two birds with one stone and combine the vetting routine at a journal with replication in the classroom for quantitative training.

When Should We Replicate?

When should studies be subject to replication assessments? One could imagine that replication of empirical results could become standard practice in the review process, where every study submitted to a journal would be scrutinized before an acceptance/rejection decision is made. However, this would add substantial effort

to the review process and is thus difficult to implement. Instead, the contributions to this forum suggest a number of alternatives. Two solutions are proposed. The “prepublication” approach suggests replicating every article that has successfully passed the peer review process. Thus, replication becomes a default quality check *before* publication, which ensures that articles not meeting even basic replication standards are not published. In other proposed approaches, replication is done *after* publication of an article, as, for example, in Janz’s (2016) replication courses or in Dafoe’s (2014) audit board proposal. Here, the assumption is that the chance that an article will be replicated serves as a deterrent for authors to publish nonreplicable findings.

Where Should Replications Be Published?

Ultimately, the results of our replication efforts should be communicated to the scientific community, so we need to ask ourselves how this should be done. Again, among the proposed solutions, there is considerable disagreement. The least demanding approach when it comes to publication is “prepublication,” where no separate publication comes out of a replication attempt. Instead, replication is a part of the standard quality control at the journal, which every article needs to pass. The audit-board approach proposed by Dafoe (2014) relies on a single replication report published once every year, where the results of the replication attempts on randomly selected articles are presented. This report would get a predefined slot in the journal.

As an alternative to these default screening procedures with little emphasis on subsequent publication, Janz (2016) proposes to publish replication studies as individual articles. Although this may be desirable when it comes to broad dissemination of findings, several issues must be considered. First, rather than communicating a general quality assessment across many articles as Dafoe’s audit report does, the fact that replication articles must pass the review process means that there is a strong bias toward failed replications. These are, of course, interesting in themselves, but we fail to capture trends in research quality since the community never learns about successful replications. Second, several journals are reluctant to publish replication work in the first place (Ishiyama 2014). This means that replications will often appear in different outlets as the original study, which can significantly limit their visibility. Adding to this is the often extremely long publication lag (Colaresi 2016), which makes a timely scientific exchange difficult.

Conclusion

The summary provided in previous sections reveals that there is considerable agreement among the contributors when it comes to the motivation and need for replication. To this end, the contributors suggest different ways to increase transparency both for qualitative and quantitative research. For the former, techniques are proposed that make assumptions and evidence used for theory building and testing more explicit. For the latter, the challenges lie elsewhere. Here, the contributors are concerned with the implementation of replication procedures that need to be imposed externally to weed out problematic studies.

However, it is unclear how this should be done. One solution could be to implement standard replication checks on the journal side, either as a mandatory preacceptance procedure (“prepublication”) or as a postacceptance replication audit. This has the advantage of eliminating problematic work completely, or—as in the replication audit—can at least provide a representative assessment of the quality of research without a bias toward failed replications. The other solution would be to encourage graduate students to perform replications as part of their methods training, with the incentive to develop these studies into journal articles.

Although this would allow a more in-depth engagement with the replicated studies, it unavoidably not only causes bias toward problematic studies but also runs the risk of having to deal with replication-reluctant journals down the road.

So what can be done? To combine the strengths of both approaches, a dual strategy seems suitable. First, journals should implement a standard data-vetting procedure as suggested by Colaresi (2016). This simple approach minimizes the replication effort since only conditionally accepted articles are screened, but it effectively eliminates simple replication failures that are unlikely to be published when they are later detected in a published piece. This default screening procedure could be combined with graduate student teaching in order to introduce students to cutting-edge research topics and methods.

Second, with the less interesting replication exercises out of the way, journals should be encouraged to be more open to comprehensive, in-depth replication studies. These studies should provide considerable added value compared to the original piece and need to pass the respective journal's peer review process. This would allow for a more active and timely scientific debate, something that is rarely seen in our discipline. Debate is an entrenched part of the scientific process, but it is not given the space it deserves in the most important outlets. At the same time, we should not overshoot the target. The current reluctance of journals to publish replication work should not lead us to create separate replication journals and workshops to remedy this problem. Redirecting replication work to separate forums would suggest that it is an entirely different business, operating in parallel to the main scientific debate. If we believe that replication is an integral part of the debate, replication studies should be discussed and presented where this debate evolves, which is in the main journals in our field.

References

- COLARESI, MICHAEL. 2016. "Prepublication, Replication: A Proposal to Efficiently Upgrade Journal Replication Standards." *International Studies Perspectives* 17: 367–78.
- DAFOE, ALLAN. 2014. "Science Deserves Better: The Imperative to Share Complete Replication Files." *PS: Political Science & Politics* 47: 60–66.
- GLEDITSCH, NILS P., CLAIRE METELITS, AND HAVARD STRAND. 2003. "Posting Your Data: Will You Be Scooped or Will You Be Famous?" *International Studies Perspectives* 4: 89–97.
- GLEDITSCH, KRISTIAN S., AND NILS B. WEIDMANN. 2012. "Richardson in the Information Age: Geographic Information Systems and Spatial Data in International Studies." *Annual Review of Political Science* 15: 461–81.
- ISHIYAMA, JOHN. 2014. "Replication, Research Transparency, and Journal Publications: Individualism, Community Models, and the Future of Replication Studies." *PS: Political Science & Politics* 47: 78–83.
- JANZ, NICOLE. 2016. "Bringing the Gold Standard into the Classroom: Replication in University Teaching." *International Studies Perspectives* 17: 392–407.
- KATSANIDOU, ALEXIA, LAURENCE HORTON, AND UME JENSEN. 2016. "Data Policies, Data Management and the Quality of Academic Writing." *International Studies Perspectives* 17: 379–91.
- KING, GARY. 2003. "The Future of Replication." *International Studies Perspectives* 4: 100–05.
- KING, GARY, ROBERT O. KEOHANE, AND SIDNEY VERBA. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- MARTEL GARCÍA, FERNANDO. 2016. "Replication and the Manufacture of Scientific Inferences: A Formal Approach." *International Studies Perspectives* 17: 408–25.
- PIERSKALLA, JAN H., AND FLORIAN M. HOLLENBACH. 2013. "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa." *American Political Science Review* 107: 207–24.
- TUCKER, TODD. 2016. "Grounded Theory Generation: A Tool for Transparent Concept Development." *International Studies Perspectives* 17: 426–38.