

# Replicate Others as You Would Like to Be Replicated Yourself

Nicole Janz, *University of Nottingham*

Jeremy Freese, *Stanford University*

Merton (1973 [1942]) famously presented “organized skepticism” as a necessary normative condition for effective science. To succeed as a self-correcting enterprise, scientific communities cannot wall off any part of themselves from reevaluation and potential revision. One revelation of the open science movement has been how much the conventional “closed-science” practices prevailing in much of political science and elsewhere undermine the possibility for effective critical scrutiny (Elman, Kapiszewski, and Lupia 2018).

Replication projects revisit existing findings and, as such, serve as the “acme” of organized skepticism (King 1995). Replications are recognized as fundamental for the scientific enterprise in principle, but they also lead to replication projects often being discouraged and fraught in practice. In a sense, replications are deliberately not original and not pathbreaking, which diminishes the enthusiasm among journal editors to publish them. Being subject to a replication project, meanwhile, often is regarded less as flattering than as something to fear.

Replication projects are thus both necessary and intrinsically delicate endeavors. The same concerns about a “crisis of credibility” that have provoked other developments in open science also underscore that political science needs more replications. Achieving a research culture in which replications become more accepted and human errors are normalized involves cultivating principles by which replications can be undertaken in a maximally constructive way. This way, they become publishable (for the replicators) and there is nothing to fear (for the original authors).

This article presents some principles on constructive ways to conduct replications. Following the style of the Transparency and Openness Promotion (TOP) guidelines for open science (Nosek et al. 2015), we summarize recommendations as a series of tiers from what is commonly done (Level I), what would be better (Level II), and what would be better still (Level III) (table 1). The aspiration of our constructive replication recommendations is to help fields move toward a research culture in which self-correction is welcomed, honest mistakes are normalized, and different interpretations of results are recognized as a routine outcome of the process. Changing culture is always difficult, of course, but

conducting projects in line with ideals and encouraging ideals in others are available to researchers for contributing to an improved culture that is closer to reality.

## CLARIFYING PURPOSE

Researchers should be clear about why an existing finding has been selected for a replication project. When projects do not explain why a particular study was selected for reexamination, original authors may feel that they are being personally attacked. Journal editors may be puzzled about the stakes in revisiting this specific finding as opposed to the many other published findings that no one else has tried to replicate. Given how many published papers are never cited, simple skepticism alone is a weak justification for all of the effort that a conscientious replication project entails.

A clear explanation of why a study’s claim was chosen to be revisited also helps readers to understand the value of the undertaking. The best rationales for undertaking a replication project connect its implications to the broader influence of the study in question or to broader debates of which the original study is a part.

Also, researchers who call their studies “replications” should not rely on that term to convey the aim of their study or the logical relationship between their findings and that of the original study. “Replication” is used to encompass various activities. The common thread is that instead of adopting published findings as *premises* on which to build, replication projects interrogate the sturdiness of those findings. At one end, replication borders *verification*: the effort to determine whether using the same methods on the same data produces the published result (called “duplication” by King 1995, 451, and currently often discussed as “reproducibility”). At the other, replication borders *extension*: the effort to elaborate findings or increase their scope.

Recognizing the variation among replication activities is especially important because of the differing implications about what divergent results between a replication project and an original study imply. The reason for a failed effort to duplicate another’s findings using the same materials and methods is almost always knowable, even if it can be frustratingly difficult to figure out. Whenever replication introduces

new data, the interpretation of differences becomes more complicated. Reasons for divergent results are often mysterious without running more studies and using more data; even then, they commonly elude decisive explanation. In these cases, it is too simplistic to believe that we can settle the “truth” of a claim between only two studies (i.e., an original and a replication). In fact, a good replication reflects on typical robustness challenges in particular fields; ideally, it should be framed as such.

strong for what divergent results imply, and we encourage replicators to choose their wording carefully.

**TRANSPARENCY**

We are enthusiasts for transparent research practices in general, and the many good arguments for research transparency apply straightforwardly to replications as well. In addition, given the direct and potentially delicate relationship to others’ work, those conducting replications are obliged to make the

*Achieving a research culture in which replications become more accepted and human errors are normalized involves cultivating principles by which replications can be undertaken in a maximally constructive way.*

Given the range of activities that may be called “replication,” the most constructive way for investigators to proceed is to be clear about what it aims to do and what it considers to be the logical relationship between its findings and those of the original study. This is especially important when results diverge. Even when a well-conducted “failed” replication causes us to reassess a previous finding, we should not automatically interpret it as discrediting an original author’s diligence, expertise, and choice of methods. The connotation of “failed replication” therefore is usually too

details of their work as open to scrutiny as possible. Data sharing of a replication (i.e., duplication) study should include the data, information about the collection and processing of the data, codebook, and software code, which should be uploaded to an online repository. Best practice would entail someone cross-checking that the replication study’s data and code are running correctly.

In some cases, it can be useful to preregister a replication study, especially when experiments or other projects that involve collecting new data are being replicated. If a *full*

Table 1

**Recommendations for a Constructive Replication Study**

	<b>Level I (Room for Improvement)</b>	<b>Level II (Better)</b>	<b>Level III (Best Practice)</b>
<b>Clarity and Purpose of Study Selection</b>	Unclear why original study was selected, or which claim exactly is being assessed.	Clear explanation of why original study was chosen, but the focus lies on only one paper’s claims for verification.	Clear explanation of why original study was chosen, and the replication has a wider purpose and examines several claims in the literature.
<b>Transparency</b>	No data sharing.	Replication data and code available at online repository. <i>In some cases:</i> Replication study was preregistered to avoid file-drawer problem.	Replication data and code available at online repository; cross-check of the replication study’s data and code was conducted independently. <i>In some cases:</i> Full analysis plan was preregistered to avoid p-hacking or other accusations.
<b>Involving Original Author(s)</b>	Original author(s) not informed.	Original author(s) informed of the replication attempt but not given the opportunity to comment on the replication procedure or the results.	Original author(s) informed of replication attempt and invited to comment on the research design, procedures, and final result throughout the various stages of the replication project.
<b>Results Write-Up</b>	Binary judgment in which the replication is framed as failure versus success. Only superficial interpretation of divergent results. Harsh, personal, and confrontational language.	Nonbinary judgment and detailed interpretation of differences between original and replication. Professional and nonpersonal language.	Nonbinary judgment, detailed interpretation of differences between original and replication. Replication study acknowledges that multiple interpretations are possible and engages deeply with substantive literature. Replication study points to lessons learned for methodological and theoretical improvement in the wider field. Professional, courteous, and collaborative language.

pre-analysis plan is preregistered, replicators can describe their decision-making process and planning and can avoid being accused of p-hacking for the purpose of error hunting. Some replication projects start as extensions, only reverting to being replication projects when problems are encountered that lead a researcher to question whether an earlier study should be considered more closely. For this reason (and others), we are wary of any dictum in which replication projects without

implementation errors” (Imai 2005, 283). The original authors then replied that the replication contains “statistical, computational, and reporting errors that invalidate its conclusions” (Gerber and Green 2005, 301), stating that “none of the key substantive or methodological claims of Professor Imai’s essay survives scrutiny.” This replication chain shows that the language—although professional—can be relatively strong and become personal between original authors and replicators.

### *In most cases, a binary judgment—that is, replication has failed versus succeeded—is unhelpful.*

preregistration are considered substandard. However, the capacity of preregistration to strengthen a replication project is obvious and we strongly urge investigators to consider it when embarking on a project.

Preregistration also provides an opportunity for investigators to inform authors of the original study about the replication project and elicit feedback about the planned research. Researchers should never need “permission” from a study’s authors to undertake a replication project. However, informing the original authors early on bears great potential for replicators to benefit from their insights and to set a collaborative tone. If original authors learn of a replication project only after it is accepted for publication, they might feel “ambushed,” and it is difficult to see any scientific purpose being served by delaying their potential input. If original authors are not involved before a replication project is conducted, we think it is better at least to provide them a copy of the replication before journal submission, especially if the results challenge the original finding.

#### **HOW REPLICATORS AND ORIGINAL AUTHORS (SHOULD) TALK TO ONE ANOTHER**

The few replications published in political science are relatively harsh in their wording, which may be related to surviving the peer-review process. For illustration, we describe a few “replication chains” of scholarly communication<sup>1</sup>: the chain progresses from the original study, to a replication attempt, to a further response from the original author on the replication. Original authors generally tend to defend their earlier paper by pointing to flaws in the replication.

A different replication chain uses slightly more positive wording. An original article on citizens’ political tolerance (Peffley, Knigge, and Hurwitz 2001a) was replicated with the statement, “We regret to say that we found some significant differences when attempting to replicate” the study (Miller et al. 2001, 407). The replicators discussed reasons for the differences in the results and showed how they double-checked their own procedures: “[w]e wanted to give the original analyses the benefit of the doubt” (Miller et al., 408). The replication study concluded that the original authors “made a simple coding mistake” and that “these analysis errors are not significant enough to dismiss this article totally, [even though] they are troublesome” (Miller et al. 2001, 409). In response, the original authors commented that the replication is “based on a fundamentally flawed analysis,” that the replicators’ criticisms “suffer from a limited understanding of existing theory and research in the area,” and that the alarm raised by the replicators is “seriously exaggerated” (Peffley, Knigge, and Hurwitz 2001b, 421–22).

What is particularly noteworthy about these scholarly exchanges is that the replicators as well as the original authors used a detailed comparison of differences of results, which reflects good practice. However, they varied in their interpretation of the importance of these differences, and the interpretation can easily take a personal, accusatory tone.

In most cases, a binary judgment—that is, replication has failed versus succeeded—is unhelpful. Instead, a constructive replication interprets diverging results step by step and takes

### *We need to develop a research culture in which authors view their study being chosen for replication as a sign of interest and importance rather than something to be feared.*

For example, a highly cited randomized field experiment published in the *American Political Science Review* showed that voter turnout increased after personal canvassing but not telephone calls (Gerber and Green 2000). A study later reassessed the results and reported that “Gerber and Green’s negative finding is caused by inadvertent deviations from their stated experimental protocol,” pointing to “systematic patterns of

care to be clear that it is discussing the “study,” not the “author.” Often, the best replications accept that multiple interpretations are possible, presenting lessons learned for methodological and theoretical improvement in the wider field. Professional, courteous, and collaborative language can communicate these improvements without alienating the original author.

### A CONSTRUCTIVE REPLICATION

The expertise and professionalism of replicators should be reflected in their careful and transparent planning of a replication study and in their professional wording. As a positive example, Busby and Druckman (2018, 5) reexamined previous research about irrelevant events, a field that is concerned with influence such as the weather or sporting events on political attitudes. The replicators noted:

To be clear, this is not a critique of existing papers, which faithfully report careful studies that establish the existence of irrelevant event effects... Rather, replication with a different event, sample, and time is a way to move the literature forward to assess robustness and the conditions under which irrelevant event effects occur.

The replicators also acknowledge up front that the original authors did not, in fact, claim generalizability. The replicators then avoid a binary judgment, such as that the previous study failed to replicate, but instead state clearly which results replicated and which did not. The replicators also emphasize that their replication result should “not be taken as definitive evidence that the extant literature overstates the extent of irrelevant events; yet, it serves as a (cautionary) prompt to the next generation of work” (Busby and Druckman 2018, 8). There is only one shortcoming of our illustrative example: Busby and Druckman (2018) were replicating their own study (Busby, Druckman, and Fredendall 2017). Does this mean it is irrelevant? On the contrary, this replication study serves exactly the point we are making.

Our golden rule is: *Replicate others as you would like to be replicated yourself*. Imagine that you look back on one of your previous studies and you believe that something is missing or that a follow-up and cross-check is in order to advance the literature on the topic. Then plan and write your replication of another’s study exactly as you would have done it for yourself.

### CONCLUSION

This article focuses on what replicators can do to provide the most constructive replications. However, the way in which original authors react also plays a role. We need to develop a research culture in which authors view their study being chosen for replication as a sign of interest and importance rather than something to be feared. A survey recently showed that when original authors responded positively to a replication of their work, conducted a follow-up study, or published a self-correction, they were seen as highly ethical and competent by other researchers and the public (Ebersole, Axt, and Nosek 2016). Similarly, being open to student replications as part of their research methods training can earn original authors praise because they are effectively supporting education of future cohorts (Ball and Medeiros 2012; Janz 2016). Analogous to our golden rule for replicators, we therefore recommend to original authors: *Talk to your replicators as you would want them*

*to talk to you*. When both replicators and original authors adhere to the highest transparency and communication standards, replications will become more acceptable and recognized as a welcome way to add knowledge to the field of political science.

### ACKNOWLEDGMENTS

We thank Richard Ball, Nate Breznau, Chris Chartier, Gary King, Arthur Lupia, Florian Markowitz, Norm Medeiros, David Mellor, Brian Nosek, and Joe Young for their comments, as well as those of the anonymous reviewers. An earlier version of this article was presented at the MZES Open Social Science Conference 2019 in Mannheim. ■

### NOTE

1. Several blog posts on “replication chains” and communication between original authors and replicators are published on the *Political Science Replication Blog*, available at <https://politicalsciencereplication.wordpress.com/category/replication-chains> (accessed January 2, 2019).

### REFERENCES

- Ball, Richard, and Norm Medeiros. 2012. “Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis.” *Journal of Economic Education* 43 (2): 182–89.
- Busby, Ethan C., and James N. Druckman. 2018. “Football and Public Opinion: A Partial Replication and Extension.” *Journal of Experimental Political Science* 5 (1): 4–10.
- Busby, Ethan C., James N. Druckman, and Alexandria Fredendall. 2017. “The Political Relevance of Irrelevant Events.” *Journal of Politics* 79 (1): 346–50.
- Ebersole, Charles R., Jordan R. Axt, and Brian A. Nosek. 2016. “Scientists’ Reputations Are Based on Getting It Right, Not Being Right.” *PLoS Biol* 14 (5): e1002460.
- Elman, Colin, Diana Kapiszewski, and Arthur Lupia. 2018. “Transparent Social Inquiry: Implications for Political Science.” *Annual Review of Political Science* 21:29–47.
- Gerber, Alan S., and Donald P. Green. 2000. “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment.” *American Political Science Review* 94 (3): 653–63.
- Gerber, Alan S., and Donald P. Green. 2005. “Correction to Gerber and Green (2000), Replication of Disputed Findings, and Reply to Imai (2005).” *American Political Science Review* 99 (2): 301–13.
- Imai, Kosuke. 2005. “Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments.” *American Political Science Review* 99 (2): 283–300.
- Janz, Nicole. 2016. “Bringing the Gold Standard into the Classroom: Replication in University Teaching.” *International Studies Perspectives* 17 (4): 392–407.
- King, Gary. 1995. “Replication, Replication.” *PS: Political Science & Politics* 28 (3): 444–52.
- Merton, Robert K. 1973 [1942]. “The Normative Structure of Science.” In *The Sociology of Science: Theoretical and Empirical Investigations*, ed. Robert K. Merton, 267–80. Chicago: University of Chicago Press.
- Miller, Arthur, Tor Wynn, Phil Ullrich, and Mollie Marti. 2001. “Concept and Measurement Artifact in Multiple Values and Value Conflict Models.” *Political Research Quarterly* 54 (2): 407–19.
- Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck, et al. 2015. “Promoting an Open Research Culture.” *Science* 348 (6242): 1422–25.
- Peffley, Mark, Pia Knigge, and Jon Hurwitz. 2001a. “A Multiple Values Model of Political Tolerance.” *Political Research Quarterly* 54 (2): 379–406.
- Peffley, Mark, Pia Knigge, and Jon Hurwitz. 2001b. “A Reply to Miller et al.: Replication Made Simple.” *Political Research Quarterly* 54 (2): 421–29.