On Data 'Janitor Work' in Political Science:
The Case of Thermostatic Support for Democracy

Memo to Editor and Reviewers

We want to thank the Editor and the reviewers for their helpful comments and suggestions. We have made specific revisions corresponding to your constructive comments and suggestions for improvement. Here we set out a list of the specific points raised in the reviews and our responses to them below, roughly in the order they appear in the text:

1. **Identification of "data-entry errors"**. Both the editor and two reviewers raised concerns about "data-entry errors." Editor and Reviewer 2 (R2) are skeptical about to what extent the issue we identified can be entirely attributed to "data-entry errors" rather than other researcher behaviors. We agree with this concern and add a paragraph on p.2 to address this issue specifically. In particular, we point out that it is impossible to tell with complete certainty the exact reason for the issues we identify, since the authors' intentions are not detectable from the outcomes. We prefer, as a rule, to attribute these issues to simple errors rather than to suspect authors of p-hacking or related types of intentional manipulations. Nor is the point we raise here to criticize this particular result, but to spotlight "typical robustness challenges"—as the work of Janz and Freese (2021, 306) suggests a good replication should do—and illuminate how idiosyncratic manual data entry processes can cause problems.

Reviewer 1 (R1) also reminded us to exercise caution when discussing the implications, considering that our research begins with the replication of a single study. We appreciate this prudence and have incorporated it into the manuscript (pp. 2 and 12). On one hand, we explicitly advise readers to recognize that our suggestions are based on an experience with a specific study and should not be seen as a universal remedy for data processing issues across all types of studies. On the other hand, we hope readers will view these recommendations as an effort to foster more reliable and transparent research. In this sense, the insights we provide can extend beyond the confines of the specific case, topic, or even discipline.

2. **Concern of promoting the "automated data entry"**. R1 has raised an important concern regarding our promotion of "automated data entry" based on a single-case study. We take this concern seriously and agree with R1 that the automation of data entry may not be suitable for all types of studies. It is also important to note that our suggestion does not imply the superiority of machines over humans in all scenarios. We address these points the discussion of our first recommendation on p. 10, explicitly highlighting the specific problems that data-entry automation aims to solve and those that it does not. We would also like to express our gratitude to R1 for suggesting Weidmann (2023) which offers a more systematic survey of the strengths of data-entry automation. We have acknowledged the existence of these studies in our manuscript (pp. 10-11), providing readers with additional resources to explore.

3. **Connection with the broad Open-Science literature**. We appreciate the Editor and reviewers for urging us to engage with the broader literature on open science and "data analytics that are robust and trusted" (DART). We have established the connections with this literature at the beginning of the manuscript (p. 2) and further elaborated on it towards the end (pp. 11-12).

We specify that this research contributes to the literature by emphasizing the significance of focusing on data preprocessing, a crucial yet often overlooked phase and illustrating its consequences for researchers with a real-life example (p.1–2 and footnote 1). In footnote 1, we also highlight the necessity of this study as a distinct research

endeavor, separate from the existing APSR piece that focuses on the same case. We greatly appreciate R1 for encouraging us to clarify this distinction and further underscore the contribution of our research.

In the final section, we show the underlying logic shared by our research and the ongoing efforts in and out of political science to promote transparent and reliable research. We thank both reviewers for suggesting branches of relevant literature.

Thank you once more for the opportunity to make these revisions. We think the paper is now stronger as a result of your comments and that it will even better promote open science standards in our discipline. We hope you all agree.

Reference

Janz, Nicole, and Jeremy Freese. 2021. "Replicate Others as You Would Like to Be Replicated Yourself." *PS: Political Science & Politics* 54 (2): 305–8. https://doi.org/10.1017/S1049096520000943.

Weidmann, Nils B. 2023. *Data Management for Social Scientists: From Files to Databases.* Methodological Tools in the Social Sciences. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108990424.