Revisiting the Evidence of Thermostatic Response to Democratic Change: Degrees of Democratic Support or Degrees of Researcher Freedom?*

Yue Hu   Yuehong Cassandra Tai   Frederick Solt

Abstract

'Janitor work'—getting data into a format appropriate for analysis—has grown increasingly important as political science research has come to depend on data drawn from hundreds of sources. One tempting solution is to simply enter data by hand, but this approach raises serious risks of data-entry error, a difficult-to-catch problem with the potential to fatally undermine our conclusions. Underscoring these points, we identify data-entry errors in a prominent recent article, Claassen's 2020 study examining how changes in democracy influence public support for democracy, and show that when these errors are corrected, its models provide no support for its conclusions. Researchers should refrain from hand-entering data as much as possible, and we offer additional suggestions for avoiding errors.

Intro

Researcher degrees of freedom, reproducibility crisis

Coding Rules and Democratic Support

With democracy under threat in countries around the world, how the public reacts to democratic erosion is a crucial question. According to a classic and still vibrant literature, growing experience with democratic governance helps generate robust public support for democracy (see, e.g., Lipset 1959a; Welzel 2013; Wuttke, Gavras, and Schoen 2022). Claassen (2020a)

*Corresponding author: yuehong-tai@uiowa.edu. Current version: October 30, 2023. Replication materials and complete revision history may be found at https://github.com/fsolt/dem_mood.

argues instead that democratic support behaves thermostatically: that increases in democracy yield an authoritarian backlash in the public, while democratic backsliding prompts the public to rally to democracy's cause.

The evidence it offers in support of this latter argument takes advantage of recent advances in modeling public opinion as a latent variable to measure democratic support. This approach provides estimates of the paper's dependent variable for over one hundred countries for up to nearly three decades, constituting a much larger evidentiary base than any previous study. These latent-variable estimates, in turn, were based on thousands of nationally aggregated responses to dozens of different questions from cross-national survey projects (Claassen 2020a, 40). Two pieces of data were collected for each distinct survey item in each country and year it was asked: the number of respondents to give a democracy-supporting response—defined, for ordinal responses, as those above the median value of the scale (Claassen 2020a, Appendix 1.3)—and the total number of respondents to whom the question was posed. Each of these 7,538 pieces of source data is recorded in a spreadsheet.[1]

We re-collected all of the source data for the publication from the original surveys. We identified the variables of the survey items used by the article within each survey dataset, and then we used an automated process to collect the needed data from the survey datasets while avoiding data-entry errors (see Solt, Hu, and Tai 2018). In Figure **??**, we compare the percentage of respondents to give a democracy-supporting response in the publication spreadsheet with the percentage we found using our automated process of wrangling these same data.

But for the remaining observations, the difference was often substantial due to data-entry errors in the publication data. For example, the Asia Barometer asked respondents in 35 country-years to indicate whether they thought "a democratic political system" would be very good, fairly good, or bad for their country. According to the study's coding rules (see Claassen 2020a, Appendix 1.3), only answers above the median of the response categories should be considered as democracy supporting, yet in this case the lukewarm intermediate

[1]The article's replication materials include only the latent variable estimates without the original survey aggregates that serve as their source data (see Claassen 2020c). Fortunately, however, the spreadsheet recording these original source data is included in the replication materials for a companion piece that employed the identical estimates (see Claassen 2020b).

category was coded as supporting democracy as well.[2]

Similarly, the four waves of the Asian Barometer included the following item: "Here is a similar scale of 1 to 10 measuring the extent to which people think democracy is suitable for our country. If 1 means that democracy is completely unsuitable for [name of country] today and 10 means that it is completely suitable, where would you place our country today?" In accordance with the coding rules of the study, responses of 6 through 10 are considered democracy supporting, and that is how the first, third, and fourth waves of the survey are coded. For the second wave, however, 5 was erroneously also included among the democracy-supporting responses.

A third example comes from the Pew Global Attitudes surveys' four-point item asking about the importance of living in a country with regular and fair contested elections: the question wording is "How important is it to you to live in a country where honest elections are held regularly with a choice of at least two political parties? Is it very important, somewhat important, not too important or not important at all?" In this case, rather than including respondents who gave both responses above the median—"very important" and "somewhat important"—only those respondents who answered "very important" were entered as supporting democracy. This error caused substantial underreporting of the extent of democratic support in 91 country-years.

According to the study's coding rules, refusing to answer is equivalent to answering in a fashion not supporting democracy (see Claassen 2020a, Appendix 1.3).

Finally, although not depicted on this plot, data-entry errors were also evident in the variable recording the year in which a survey was conducted: these typically reflected differences between the nominal year of a survey wave and when the survey was actually in the field in a particular country.

[2]Although this may be interpreted as an exercise of researcher judgment as to what constitutes a democracy-supporting response rather than a data-entry error, examination of similar answers to similar questions shows that similarly lukewarm responses at and below the median response category (e.g., in the Arab Barometer, that democracy was "somewhat appropriate" for the country) were coded as not supportive.

Consequences for Inference

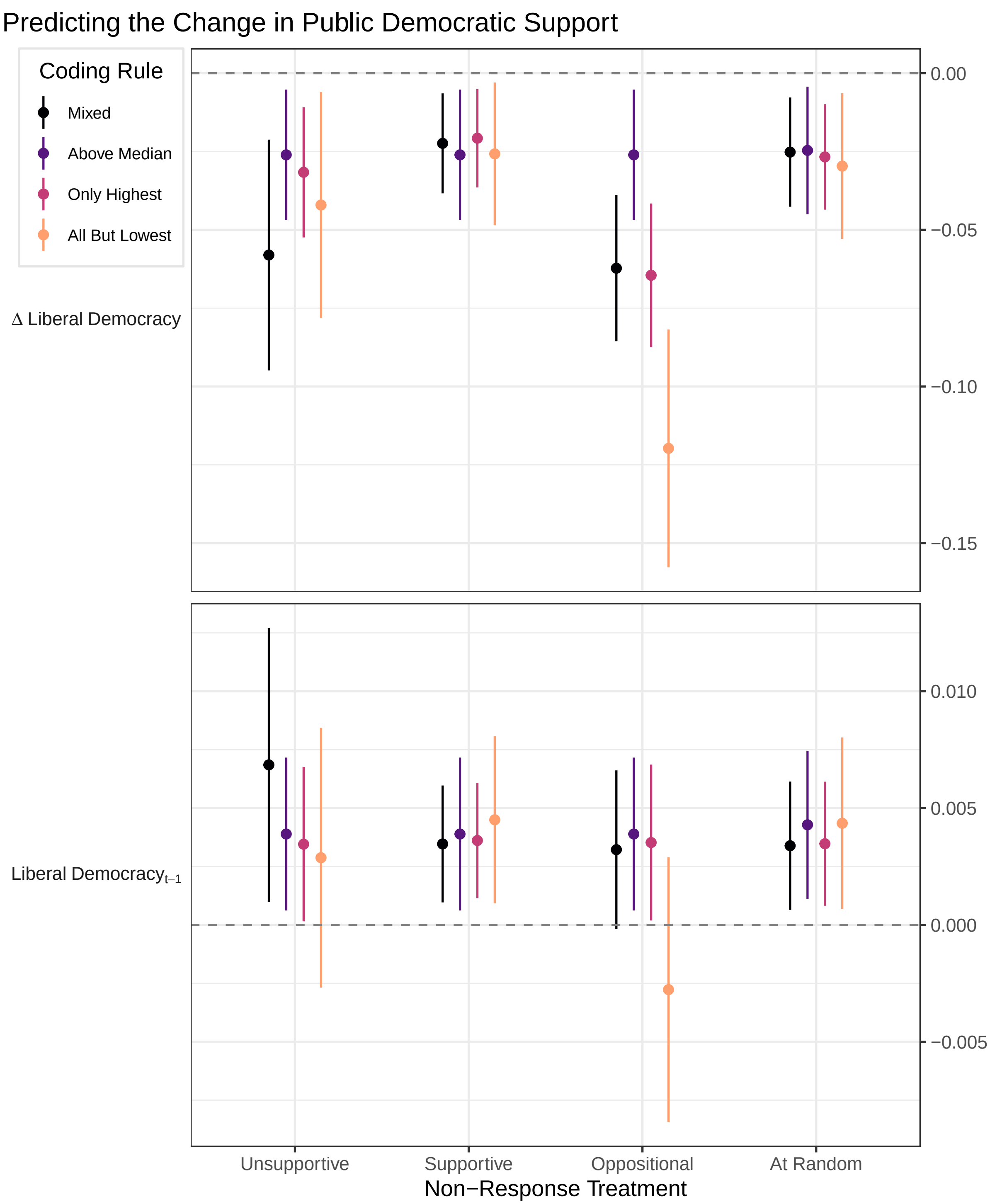
Data-entry errors of this sort can yield erroneous conclusions. After generating the latent variable of democratic support with the corrections to the errors described above, we replicated each of the models presented in Claassen (2020a) exactly using both the original and the new version of the latent variable. The results using the corrected data reveal some support for the classic argument that democracy generates its own demand through the long-run processes of socialization and learning and none at all for a thermostatic relationship.

Figure 1 is a "small multiple" plot (see Solt and Hu 2015) showing the results of replicating Model 1.1, the principal model of Claassen (2020a, 47), with each of the sixteen combinations of coding rule and treatment of survey non-response. In the top panel, the dots represent point estimates for the coefficients for change in liberal democracy; in the bottom panel, they depict the coefficients for lagged level of liberal democracy. In both panels, the whiskers show the associated 95% confidence intervals. Each coding rule is represented by a different color, while the four non-response treatments are shown in separate clusters from left to right.[3]

Models 1.1 through 1.4, which replicate those presented in Table 1 of Claassen (2020a, 47), examine the effects of overall liberal democracy using error-correction models and first-difference models. As Claassen (2020a, 46) notes, the thermostatic theory predicts that the estimated coefficient of the change in liberal democracy will be negative, while the classic theory suggests that lagged levels of liberal democracy will be positive. When using the original publication data with their data-entry errors, we replicate the results of the article exactly: the coefficients estimated for the change in liberal democracy are large, negative, and statistically significant across all four models, just as the thermostatic theory predicts. The positive and statistically significant result for the lagged level of liberal democracy found in Model 1.1—supporting the classic theory—disappears when corruption is taken into account in Model 1.2.

When the data-entry errors are corrected, however, the results for these models suggest a very different set of conclusions. The standard errors shrink across the board, indicating that the models are better estimated in the corrected data. The positive and statistically significant result for the lagged level of liberal democracy remains in Model 1.1, and the

[3]The full results for Model 1.1 can be found in the online Supplementary Materials.



Notes: Replications of Claassen (2020, 47), Table 1, Model 1.1. The mixed coding rule employed in Claassen (2020) along with that work's assumption that non−responses indicate a lack of support for democracy yields a larger negative point estimate of the coefficient for change in liberal democracy than most other combinations and a larger point estimate of the coefficient for the lagged level of liberal democracy than all other combinations. In error−correction models like these, both coefficients must be interpreted together; see Figure 2.

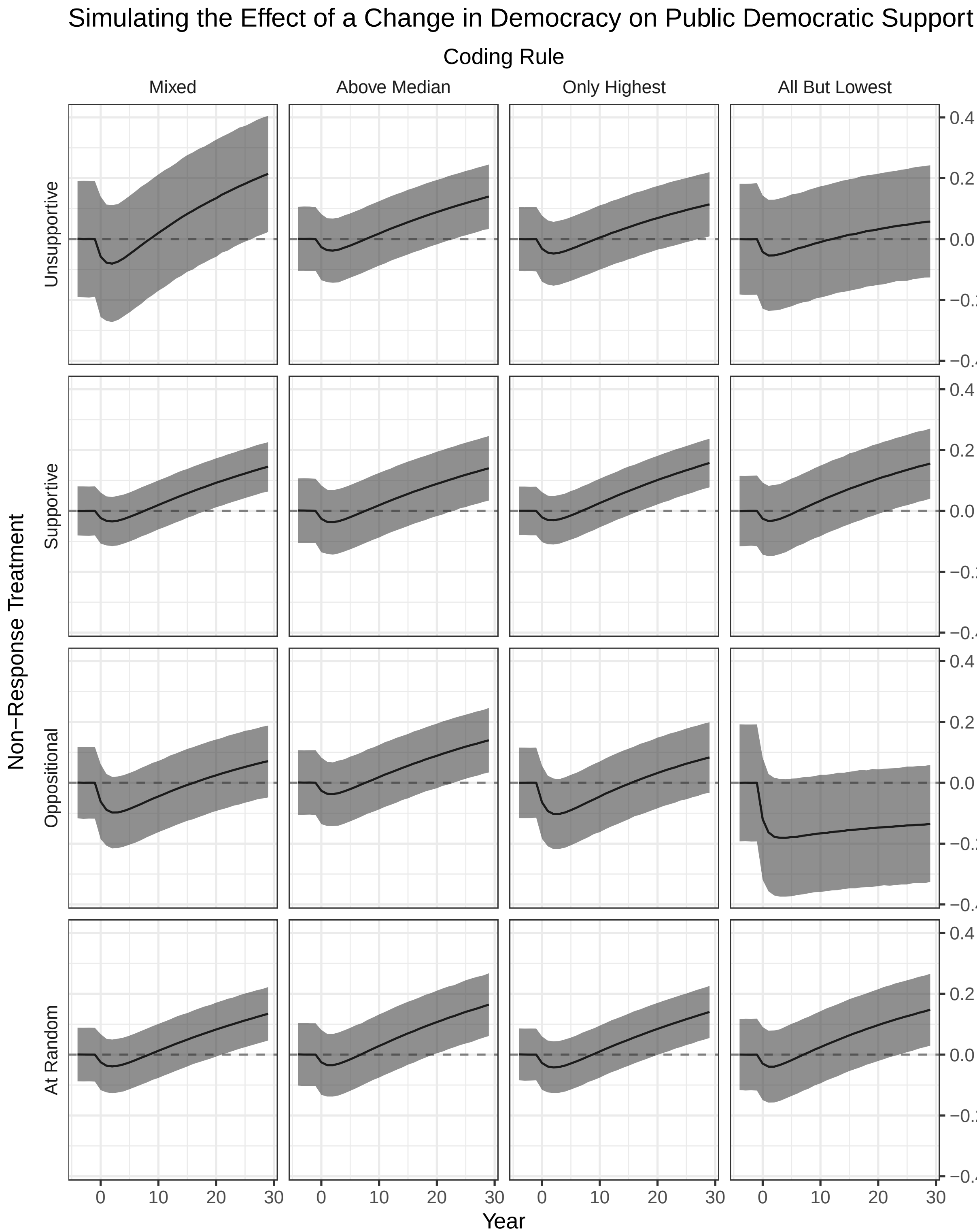
Figure 1: The Effects of Democracy on the Change in Public Support

estimate is only slightly smaller than in the publication data. When corruption is added in Model 1.2, or when the focus shifts from overall liberal democracy to its minoritarian component in Models 2.1 and 2.2, this estimate actually grows slightly larger when using the corrected data rather than the publication data, although it still fails to reach statistical significance. On the other hand, the estimates for the change in liberal democracy that provided support for the thermostatic theory are much smaller in all of the models. Models 2.1 through 2.4, which break liberal democracy into its electoral and minoritarian components, similarly yield smaller coefficients for the change in democracy than those found using the publication data.

Like the coefficients of constitutive terms of multiplicative interactions (see, e.g., Brambor, Clark, and Golder 2006), the coefficients of error-correction models cannot be interpreted separately (see Williams and Whitten 2012). Figure 2 is similar to Claassen's (2020a, 48) Figure 5. It depicts simulated effects, in differences not levels for ease of interpretation, of a one standard deviation increase in democracy on the public's support for democracy using the sixteen sets of regression coefficients presented in Figure **??** with the four different coding rules appearing the columns and the four different missing data treatments in the rows.

In the upper left pane, the combination of mixed coding rule and treating survey nonresponses as indicating a lack of support for democracy matches that employed in Claassen (2020a). The initial drop and slow recovery in the mean of these simulations of public democratic support was the evidence presented for the "thermostatic response of public opinion" and the claim that "little evidence that democracy generates its own demand" (Claassen 2020a, 48). But the other panes, with a few exceptions, show smaller dips, quicker recoveries, and continued increases; these findings lead to very different conclusions. Indeed, most of these alternate analyses—eleven of fifteen—show statistically significant increases in democratic support within three decades, the sort of generational change predicted by the classic theory since Lipset (1959b). None of them show statistically significant declines that would lend credence to the argument that democratic support responds thermostatically.[4]

[4] Replications of the article's other models can be found in the online Supplementary Materials.



Notes: Simulated effects are estimated using coefficients from the models presented in Figure 1. The solid lines indicate the mean simulated effect; the shaded regions indicate the 95% confidence intervals of these effects.

Figure 2: Simulated Effects of Democracy on Changes in Public Democratic Support

Discussion

That support for democracy is different from the other aspects of public opinion that exhibit thermostatic responses should not be surprising: the theory's mechanism does not apply to democratic support. As originally proposed, the theory demanded a level of political knowledge that the public is well understood to not hold, and as recently re-elaborated it requires political parties to debate the issue as so provide the public with cues as to what is going on (Atkinson et al. 2021, 5–6). But political parties that work to roll back democracy rarely if ever explicitly argue for that outcome. Instead, such parties insist that their actions are necessary to protect democracy from pernicious external influences or are needed to reform democracy in ways that will better serve national interests. And when these parties manage to erode democracy by restricting the freedoms of speech, press, and association, opposing parties and their counterarguments grow less and less likely to even reach the public. Without open and vigorous debate on the issue, any thermostatic response in democratic support to democratic erosion breaks down.

The analysis above reveals that data-entry errors are an especially pernicious threat to the credibility of our results. The threat is a subtle one that is not easily detected. To discern it requires close scrutiny of every manual entry; merely examining the data and their distribution will uncover few errors (Barchard and Pace 2011a, 1837–38). Although failure to find support for a research hypothesis may prompt us to undertake a such a close review, an analysis that yields statistical significance is unlikely to trigger what will likely be, as in the above example, a time-consuming and difficult effort (see Gelman and Loken 2014, 464).

This leads us to recommend the following steps to reduce possible data-entry errors. First, **automate data entry**: we suggest researchers to consider reducing reliance on manual data entry and increase the extent to which data wrangling is performed computationally. Readers should be aware that our suggestion does *not* imply that computers are always superior to people for this purpose. Instead, automated coding should always involve sensitive human design and systematic supervision (see Breznau 2021; Grimmer 2015). Still, given the convenience of automatic data entry in documentation and reproduction, we encourage researchers to use it as much as possible instead of entering data manually to increase the efficacy and transparency of their data processing operations.[5]

In making this recommendation, we are aware that being open and transparent in this way takes effort (Engzell and Rohrer 2021). But as researchers automate more of their data entry, the chances that they can reuse their code in subsequent projects improve. In fact, many common janitor-work chores already have been packaged as open-source software that to make researchers' task more straightforward and easier.[6]

Second, **use the double-entry method**: when manual data entry cannot be avoided, each entry should be made twice. Double entry is labor intensive, but experiments have shown that it reduces error rates by thirty-fold even when done immediately after the initial collection and by the same person (Barchard and Pace 2011a, 1837). Given that data-entry errors can completely undermine the validity of our conclusions, as in the example above, double entry is worth the extra effort.

Third, **embrace teamwork**: for any project involving entering data by hand, splitting the task up among team members will reduce the risk of errors going undetected. When double entries are performed by different people, discrepancies will be noted, discussed, and resolved correctly; having two sets of eyes on complex materials like survey codebooks also increases the chances that nuances of the presentation like survey weights will be uncovered. Further, by dividing the load, teamwork also lessens the probability of errors due to fatigue arising in the first place.

Fourth, **be aware of the threat of data-entry error**: this final recommendation is especially for manuscript reviewers. If data-entry errors are invisible to the authors themselves, they are doubly so to reviewers (though if editors provided reviewers with replication materials at the time of the review it may help them to better assess the work's credibility). But the case described above nevertheless suggests a valuable heuristic: when a work's conclusions suggest that a difficult problem will be easily solved—that democratic erosion will reflexively trigger a backlash and a renewed public support for democracy, in the present instance—it warrants especially careful scrutiny.

[5] For a systematic discussion of the function of automated data processes in social science, see Weidmann (2023).

[6] For example, see `readtext` (Benoit et al. 2016) for formatting text files and `DCPOtools` (Solt, Hu, and Tai 2018) for aggregating cross-sectional time-series public-opinion surveys.

Data-entry errors are inevitable, and even following these recommendations is unlikely to eliminate them entirely. Further, the above suggestions follow closely from a specific case and, although they successfully help us identify and fix its data-entry issues, they do not constitute a panacea to cure all data-processing problems in all types of research.

Nonetheless, we also hope the readers to see the shared logic of these suggestions and the growing literature to guide political scientists to conduct more reliable and credible research. For instance, in the same vein as our first suggestion, Weidmann (2023) provides a book-length set of illustrations on how to reduce "manual point and click" tasks found in a variety of studies with the `tidy`-data framework in the R language. Kapiszewski and Karcher (2021, 288) even suggests that qualitative researchers should consider using "open-exchange format" of qualitative data analysis software to be more "transparent about the generation and analysis of data." Furthermore, we regard our efforts and recommendations as a contribution to the open science movement to produce more robust and credible research in the social sciences (see, e.g., Christensen, Freese, and Miguel 2019) and beyond (see, e.g., Barchard and Pace 2011b; Lohr 2014). With careful attention, not only can the threat of data-entry errors to our 'janitor work', our research, and our understanding of the world be minimized, but the transparency, openness, and credibility of our research can continuously grow.

References

Atkinson, Mary Layton, K. Elizabeth Coggins, James A. Stimson, and Frank R. Baumgartner. 2021. *The Dynamics of Public Opinion.* Cambridge: Cambridge University Press.

Barchard, Kimberly A., and Larry A. Pace. 2011a. "Preventing Human Error: The Impact of Data Entry Methods on Data Accuracy and Statistical Results." *Computers in Human Behavior* 27 (5): 1834–39.

———. 2011b. "Preventing Human Error: The Impact of Data Entry Methods on Data Accuracy and Statistical Results." *Computers in Human Behavior* 27 (5): 1834–39. https://www.sciencedirect.com/science/article/pii/S0747563211000707.

Benoit, Kenneth, Adam Obeng, Paul Nulty, Aki Matsuo, Kohei Watanabe, and Stefan Müller. 2016. "readtext: Import and Handling for Plain and Formatted Text Files." Available at the Comprehensive R Archive Network (CRAN).

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63–82.

Breznau, Nate. 2021. "I Saw You in the Crowd: Credibility, Reproducibility, and Meta-Utility." *PS: Political Science & Politics* 54 (2): 309–13. https://doi.org/10.1017/S1049096520000980.

Christensen, Garret, Jeremy Freese, and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Berkeley: University of California Press. https://www.google.com/books/edition/Transparent_and%3Csub%3ER%3C/sub%3Eeproducible%3Csub%3ES%3C/sub%3Eocial%3Csub%3ES%3C/sub%3Ecie/PvqVDwAAQBAJ?hl=en&gbpv=1&dq=Transparent+and+Reproducible+Social+Science+Research:+How+to+Do+Open+Science&printsec=frontcover.

Claassen, Christopher. 2020b. "Replication Data for: Does Public Support Help Democracy Survive?" https://doi.org/10.7910/DVN/HWLW0J, American Journal of Political Science Dataverse.

———. 2020a. "Replication Data for: Does Public Support Help Democracy Survive?" https://doi.org/10.7910/DVN/HWLW0J, American Journal of Political Science Dataverse.

———. 2020c. "Replication Data for: In the Mood for Democracy? Democratic Support as Thermostatic Opinion." https://doi.org/10.7910/DVN/FECIO3, American Political Science Review Dataverse.

Engzell, Per, and Julia M. Rohrer. 2021. "Improving Social Science: Lessons from the Open Science Movement." *PS: Political Science & Politics* 54 (2): 297–300. https://doi.org/10.1017/S1049096520000967.

Gelman, Andrew, and Eric Loken. 2014. "The Statistical Crisis in Science." *American Scientist* 102 (6): 460–65.

Grimmer, Justin. 2015. "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together." *PS: Political Science & Politics* 48 (1): 80–83. https://doi.org/10.1017/S1049096514001784.

Kapiszewski, Diana, and Sebastian Karcher. 2021. "Transparency in Practice in Qualitative

Research." *PS: Political Science & Politics* 54 (2): 285–91. https://doi.org/10.1017/S1049096520000955.

Lipset, Seymour Martin. 1959a. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53: 69–105.

———. 1959b. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53 (01): 69–105.

Lohr, Steve. 2014. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights." *The New York Times*, August. https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html.

Solt, Frederick, and Yue Hu. 2015. "dotwhisker: Dot-and-Whisker Plots of Regression Results." Available at the Comprehensive R Archive Network (CRAN). http://CRAN.R-project.org/package=dotwhisker.

Solt, Frederick, Yue Hu, and Yuehong 'Cassandra'Tai. 2018. "DCPOtools: Tools for Dynamic Comparative Public Opinion." https://github.com/fsolt/DCPOtools.

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.

Welzel, Christian. 2013. *Freedom Rising: Human Empowerment and the Quest for Emancipation.* Cambridge: Cambridge University Press.

Williams, Laron K., and Guy D. Whitten. 2012. "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models." *The Journal of Politics* 74 (3): 685–93. https://doi.org/10.1017/S0022381612000473.

Wuttke, Alexander, Konstantin Gavras, and Harald Schoen. 2022. "Have Europeans Grown Tired of Democracy? New Evidence from Eighteen Consolidated Democracies, 1981–2018." *British Journal of Political Science* 52 (1): 416–28.

Online Supplementary Materials

List of Tables

List of Figures

Tabular Results

Calculating Effects Via Simulation

As in Claassen (2020a, 48–50) we estimate the effects of changes in democracy on public support for democracy in the error-correction models using simulation (see Williams and Whitten 2012). All independent variables were set to the same moderate values as in Claassen (2020a) and allowed to run for 200 years, long enough for the system of equations to stabilize. The level of democracy was then increased from half a standard deviation below the mean to half a standard deviation above; then the system of equations was allowed to run for 30 more years; these three decades those are depicted in Figure **??**. Per Claassen (2020a, Supplementary Information 3) and Claassen (2020c), the uncertainty in the model was captured by taking 10,000 draws from a multivariate normal distribution with expectation being the vector of model coefficients and variance being the robust covariance matrix, $\tilde{\Theta} \sim MVN(\Theta, \Sigma)$, and adding the noise estimated in the regression standard error, $\tilde{Y}_i \sim N(X_k \tilde{\Theta}_{ki}, \sigma)$. To get first differences, the mean value of $\tilde{Y}_i$ in the year before the increase in democracy ($t = -1$) was subtracted from each $\tilde{Y}_i$, and the 0.025 and 0.975 quantiles of the first difference were used as its lower and upper confidence bounds.

First Difference Plots for Models 2.1 and 2.2