Replication package for: Surviving the screens: The problem of hidden inattentive respondents in online surveys

Scott Blatte
University of Chicago
sblatte@uchicago.edu

Brian F. Schaffner
Tufts University
brian.schaffner@tufts.edu

May 16, 2025

This document contains instructions to replicate "Surviving the screens: The problem of hidden inattentive respondents in online surveys." We use data from the Cooperative Election Survey (CES) and from a survey administered by Lucid/Cint to define population quantities for our reweighting exercise. All data files are included under the `input` folder.

Two files used in the Appendix rely on data before YouGov implements post-survey cleaning and processing. We use these to compute weights required to estimate horserace results with and without contradictors. The data is proprietary to YouGov, and we are unable to disseminate it. If one wishes to reproduce Figures B1 and B2, they may contact YouGov to request access.

Some analyses in the Appendix are based on a replication of Kuriwaki et al. (2023). The code and data required for those files is contained in its own folder and is described in Section 4.

1 Running the code

All analysis is done in `R`. Once the replication downloads, you should see four folders: `input, code, output,` and `kuriwaki_replication`. All code is in the `code` folder and can be run in two ways. First, you can use GNU Make to run the code from the command line. Make should be automatically installed on all Mac, Linux, and Unix machines and is available for Windows. To run, open a terminal, navigate to the code folder, and type and

enter `make`. This will execute all scripts and produce all outputs without any further action. Second, you can run each script individually however you normally run R scripts. The only exception is `voteredo2020_build.R`, which is called by `voteredo2022_build_analysis.R` and does not execute or produce results on its own.

2 Description of code and outputs produced

There are six code files producing all outputs (except for some Appendix figures described below).

- `inconsistents_main_analysis.R`: This is our main analysis script. Figure 1, Table 4, and Table 5 are produced automatically in the output folder. The log file produced by this script contains the numbers to create Tables 2-3 and 6-7.
- `experiment21.R`: This script produces Figures 2-3 and Appendix Table B6 which are outputted to the output folder.
- `lucid_experiment.R`: This script produces Appendix Figure 5 in the output folder. The log file produced by this script contains the numbers for Appendix Tables B8 and B9.
- `replicate_bpz.R`: This script produces Table B3 in the output folder.
- `voteredo_2022_build_analysis.R`: This script produces Tables B1 and B2 in the output folder. Note that `voteredo_2020_build.R` is called by this script. **The required data are proprietary and not included. The code is accurate but the script will not execute without the required data.**

3 Description of inputs

- `cces_2020_LVweighted.csv`: This file contains the results of the 2020 CCES with likely voter weights computed before the election. **This file is proprietary and is not included.**
- `CCES20_Common_OUTPUT_vv.csv`: This file contains data for the 2020 Cooperative Election Study common content. See `https://doi.org/10.7910/DVN/E9N6PH`.
- `CCES21_Common_OUTPUT_timing.sav`: This file contains data for the 2021 Cooperative Election Study common content with question timing data included. See `https://doi.org/10.7910/DVN/OPQOCU`.
- `CCES22_Common_OUTPUT_vv_topost.csv`: This file contains data for the 2022 Cooperative Election Study common content. See `https://doi.org/10.7910/DVN/PR4L8P`.

- `ces_2022_weighted_final.csv`: This file contains the results of the 2022 CES with weights computed before the election. **This file is proprietary and is not included.**
- `CCES22_OUTPUT_sourcevar.csv`: This file contains information from YouGov about the 2022 common content data about whether the respondent was from the YouGov panel or from an external source.
- `poll_2022_weighted_final_topost.csv`: This file contains the results of the 2021 Lucid/Cint poll.

4 Kuriwaki et al. (2023) instructions

In the Appendix, we replicate The Geography of Racially Polarized Voting: Calibrating Surveys at the District Level. We relied on their replication package and instructions, making only minor changes within certain scripts. Most scripts are identical to the original code save for the outputs, which now include both a replication and a version where we drop contradictors. This folder mirrors their repository structure except for scripts not relevant to our exercise that we have removed. To reproduce our results, run all scripts in sequential order in `build`, then the scripts in `analyze`. Scripts `07_anova_table-02.R`, `08_coalition_figure-06-table-03.R`, and `scatterplot_differences.R` produce figures in our paper. Please refer to the included README for additional instructions. To expedite replication, we include the processed data files following the build step, which can be found in the `analyze` folder. Files with the phrase "noc" are based on the no contradictors CES. Files with the phrase "replicate" are our replications of their build. Files with neither phrase are the original files included in their replication package.