Brailey, Thomas; Hepplewhite, Matthew; Moser, Scott

Working Paper

Direct Replication and Additional Sensitivity and Robustness Analyses for Frederiksen (2022): A Replication Report from the Nottingham Replication Games

I4R Discussion Paper Series, No. 28

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Brailey, Thomas; Hepplewhite, Matthew; Moser, Scott (2023) : Direct Replication and Additional Sensitivity and Robustness Analyses for Frederiksen (2022): A Replication Report from the Nottingham Replication Games, I4R Discussion Paper Series, No. 28, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/270858

INSTITUTE for REPLICATION

No. 28
I4R DISCUSSION PAPER SERIES

Direct Replication and Additional Sensitivity and Robustness Analyses for Frederiksen (2022): A Replication Report from the Nottingham Replication Games

Thomas Brailey

Matthew Hepplewhite

Scott Moser

May 2023

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 28

Direct Replication and Additional Sensitivity and Robustness for Frederiksen (2022): A Replication Report from the Nottingham Replication Games

Thomas Brailey[1], Matthew Hepplewhite[1], Scott Moser[2]

[1]Department of Politics & International Affairs, University of Oxford/UK
[2]School of Politics and International Relations, University of Nottingham/UK

MAY 2023

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Direct Replication and Additional Sensitivity and Robustness Analyses for Frederiksen (2022): A Replication Report from the Nottingham Replication Games*

Thomas Brailey
thomas.brailey@politics.ox.ac.uk

Matthew Hepplewhite
matthew.hepplewhite@politics.ox.ac.uk

Scott Moser
scott.moser@nottingham.ac.uk

April 25, 2023

Abstract

We replicate the analysis conducted by Frederiksen, 2022a. We focus on assessing the computational and robustness replicability of their work. We find that their main exhibits and supplementary analysis are replicable, both when running their original Stata replication package, and when we attempt to replicate their findings from scratch in R. We also conduct additional robustness checks by estimating additional specifications and by subsetting the dataset by the time taken by the respondent to complete the survey. We again find that their work is robust to our battery of alternative specifications.

1 Introduction

This paper presents a replication of Frederiksen, 2022a's paper 'Does Competence Make Citizens Tolerate Undemocratic Behavior?', published in Volume 116, Number 3 (2022) of the *American Political Science Review*. Frederiksen, 2022a conducts conjoint candidate ranking experiments in five countries—the United States, the United Kingdom, the Czech Republic, Mexico, and South Korea (yielding a sample with more than 14,000 respondents and more than 260,000 candidate observations). The aim of his experiments is to test the extent to which citizens are willing to tolerate anti-democratic behaviour by politicians. Specifically, he aims to test the relationship between voters' toleration for anti-democratic leaders and the perceived competence of these leaders.

This work contributes both to the substantive field of candidate choice as well as the methodological field of conjoint experiments. The author attempts to bridge two competing theories about the successes of political leaders; the powers of partisanship and polarisation, and the perceptions of competency. This paper is also part of a growing literature that relies on conjoint experiments to measure voters' preferences for candidates. There have been, to the best of our knowledge, thirty-eight text-based candidate choice experiments, which use a large, nationally representative sample, conducted specifically in Britain, the results of which are detailed in nineteen articles (Baron, Lauderdale, and Sheehy-Skeffington, 2023; Campbell and Philip Cowley, 2014a; Campbell and Philip Cowley, 2014b; Campbell and Philip Cowley, 2018; Cowley et al., 2016; Carnes and Lupu, 2016; Eggers, Vivyan, and Wagner, 2018; Frederiksen, 2022a; Sara B Hobolt and Rodon, 2020; Kevins, 2021; Magni and Reynolds, 2021; Magni and Reynolds, 2022; Martin and Blinder, 2021; Reher, 2021; Saha and Weeks, 2022; Vivyan and Wagner, 2015; Vivyan and Wagner, 2016; Vivyan,

*Word count: 5051

Wagner, et al., 2020). Their appearance, from 2014, forms part of the wider growth of experimental research in political science (Druckman and Green, 2021).

Frederiksen, 2022a finds that competence *does not* suppress the negative effect of incremental violations of democratic principles. Further, he finds that respondents in all five countries prefer undemocratic, competent candidates to democratically compliant, incompetent candidates. Specifically, candidates who score lowest on a five-point measure of competency who are considered democratic see a roughly 0.5 point increase in favourability, also measured on a five-point scale (Frederiksen, 2022a, p. 1151). Undemocratic candidates see a similar increase in support the more competent they are, those levels of support are consistently lower than support towards democratic candidates across all models. He also includes two supplementary material documents—including specifications with additional covariates, higher-order polynomials, and average marginal component estimates—which we partially reproduce in this paper.[1]

The estimation strategy used for his primary exhibits is captured in equation 1.

$$Y_{ict} = \alpha + \beta_1 Behaviour_{ict} + \beta_2 Competence_{ict} + \beta_3 Behaviour_{ict} \times Competence_{ict} + \epsilon_{ict} \quad (1)$$

Where Y_{ict} is the dependent variable, support for a hypothetical candidate c for individual i, for task t, $\beta_1 Behaviour$ is a measure of candidate behaviour, captured in the author's vignettes (see Table 1 in Frederiksen, 2022a for an example), $\beta_2 Competence$ is a measure of candidate competence, and $\beta_3 Behaviour \times Competence$ is an interaction effect of the first two terms. Standard errors are clustered at the individual respondent level. Their main specification pools all country observations and presents effects disaggregated by country. The author includes both a multiplicative model (the equation above), as well as an additive model which simply drops the β_3 interaction coefficient.

From this regression, the author calculates the marginal means of respondents' support for a given candidate for each discrete level of candidate competence. Marginal means are defined as the mean for one variable averaged across every level of the other variables (Leeper, Sara B. Hobolt, and Tilley, 2020). The other 'industry standard' estimand, the average marginal conditional effects (AMCE) (Hainmueller, Hopkins, and Yamamoto, 2014) is a different way of assessing 'effects' of particular attributes on the ranking or choice of respondents.

1.1 Situating our Contribution

The following paper was produced at the Replication Games held on 18 March 2023 at the University of Nottingham, organised by the Institute for Replication. The aim of the Institute for Replication, established by Abel Brodeur of the University of Ottawa, is to 'improve the credibility of science by systematically reproducing and replicating research findings in leading academic journals.' (https://i4replication.org/). The Nottingham event was the fourth iteration of the Replication Games format, the previous three Games having been held in Oslo, Calgary, and Toronto (https://i4replication.org/games.html). This paper contributes to the growing literature of papers that explore the replicability and robustness of published works (Gong and Hammar, 2023; Guntermann and Lenz, 2022; Engel, Huber, and Nüß, 2022; Bonander, Strand, and Jakobsson, 2023).

The rest of the paper is organized as follows. Section 2 reproduces the main finds from Frederiksen (2022a). Section 3 performs a number of additional analysis, using different specifications, different estimators, etc. to examine the robustness of the findings. Section 4 discusses overall findings and possibilities for future research. Section 5 concludes our study.

2 Reproducibility

We begin by re-running the analysis files provided as part of the paper's replication package.[2] This portion of the analysis was originally performed entirely in Stata. There are two scripts provided in the replication package. The first, a cleaning script which takes their raw data and creates the final dataset for analysis, is not reproducible. This is because the original raw data files are not provided as part of the replication

[1] The corresponding citations for Frederiksen, 2022a's Dataverse appendix and supplementary materials documents are Frederiksen, 2022b and Frederiksen, 2022c, respectively.

[2] Replication materials we used are available at `https://github.com/tjbrailey/nottingham_replication_2023`.

package. The second, an analysis script which produces all tables and figures for both the main paper and the supplementary analysis file, is fully reproducible. We encounter no coding errors or issues with the materials provided in the replication package.

2.1 Replication

We then attempted to replicate the core exhibits in the manuscript from scratch (that is, without looking at the Stata replication code) using R: A few cleaning tasks were necessary before being able to analyse the data.[3] This mainly involved converting character variables to factors or to numeric variables. To be clear, the data we conduct our replication on is pre-processed data, as we do not have access to the raw data.

The main paper only contains one exhibit, Figure 1, replicated below (Figure 1). This figure replicates exactly. The upper panels present the marginal mean of respondents' support for democratic and undemocratic candidates as competence increases, while the lower panel present the average marginal component interaction effect (AMCIE, (Hainmueller, Hopkins, and Yamamoto, 2014)) of the same variables relative to the third category of the independent variable (`cancom`, which captures candidate competence).

Both tables in Frederiksen (2022b, Appendix A), namely Tables A1 and A2, replicate exactly in terms of point estimates, statistical significance, and number of observations, both in the original Stata code and in our R replication. The replication is given in Tables 1 and 2.

The author doesn't estimate the average marginal component effect (ACME) in the main paper, though it is worth considering the implicit weighting scheme that is suggested in the distribution of attributes as described in Table 4, as well as the 'target distribution' as given in Supplemental Appendix A of Frederiksen (2022c). We discuss this further in section 3.4.

In summary, we find the replication materials in Frederiksen, 2022a to be entirely reproducible, replicable, and very clear. We now move into assessing the robustness of his work to alternative specifications, estimators, and data truncation.

3 Extensions and Robustness

In this section we extend the analysis in a number of ways to examine the robustness of the findings.

3.1 Differential Impacts by Survey Response Time

We want to test the robustness of Frederiksen, 2022a's findings. One theory that we can test is that inference may be affected by the time taken for respondents to complete the survey. For example, those who complete the survey quickly may simply be selecting the first answer for each question. Similarly, those who spend a long time completing the survey may either be distracted, or may be conducting research in order to inform their responses. We would want to be sure that 1) the majority of respondents are taking an 'appropriate' amount of time to respond to the survey, and 2) that the exclusion of those individuals at the tails of the distribution are not affecting the point estimates or statistical significance.

Figure 2 plots the distribution of the time taken for respondents to complete the survey, (somewhat arbitrarily) truncated at 5000 seconds (∼83 minutes), for each country in the dataset. From a quick glance, we can see that the distribution in the time taken to respond to the survey differs considerably by country, with South Korea being a particular outlier.[4] While this is not of particular concern—and does not affect the conclusions drawn in this paper—we believe that the presentation of summary statistics of the meta-data of the study are important for interpretation, even if confined to an appendix.

Figure 4 plots the point estimates for each coefficient's effect on the respondent's support for the candidate—the same estimating equation used to construct Figure 1—except we truncate the sample size according to the time taken for the respondent to complete the survey. For example, the gray line (upper and

[3] Because Frederiksen, 2022a does not include any raw data, the cleaning file included in the replication package is not usable. We referred to this file when preparing the variables for analysis in R, though we did not consult the analysis code during our replication in R.

[4] Concretely, South Korea has a mean of 388 seconds, compared to the pooled mean of 670 seconds. While we do not have a strong intuition on why this is, some research suggests that these differences could be a result of cultural differences (Beuthner et al., 2018)

lower 10th percentiles dropped, that is, 20% of the data are removed) removes the slowest and fastest ten percent of respondents from the regression in order to see whether these individuals are driving any of the effects that we see. Broadly speaking, the results are robust to these different data subsets. In the Mexico sample, the statistical significance of the point estimates are sensitive to the sample truncation, though is to be expected, since we are dropping large percentages of the data when we re-estimate the effects. Another way to capture robustness to respondent attentiveness is to subset the data to those who are classed as being 'attentive' in the survey response.[5] Figure 5 plots the point estimates from the regression described in equation 1, and we again see that the results are robust to the removal of inattentive respondents.

3.2 Robustness to Survey Fatigue Effects

We have also undertaken work to test the impact of 'fatigue effects', described as 'a well-documented phenomenon that occurs when survey participants become tired of the survey task and the quality of the data they provide begins to deteriorate' Ben-Nun, 2011, p. 742. Here is the logic: Frederiksen, 2022a asks respondents to complete ten choice tasks. Each of these ten questions involves the respondent rating two candidates from 1 to 5 in terms of support. For each task (1 through 10), we can calculate the difference in ratings between each candidate pair. It is likely that sometimes this difference will be due to the responding genuinely thinking both candidates are equally as good, bad, or neutral as each other, but we might expect to see the mean difference (systematically) decrease over time if respondents are in fact getting tired (that is, as they complete more tasks, they are more likely to rate the two candidates the same as it is easier to just quickly click through the survey). Figure 3 plots this relationship. Qualitatively, this figure suggests a further analysis of 'early' respondent answers versus 'later' answers. Despite the considerable drop off in attention exhibited by respondents—at least as judged by the amount of time spent on each of the ten tasks—there appears to be none-to-very-little difference in the estimated effects using only 'early replies.' Table 3 shows the results of the estimation performed in Table 2 but only using a subset of the data. Specifically, the estimates in Table 3 are based on only each respondent's first two (out of ten) replies. Comparing Table 2 (which uses respondents' answers to all ten prompts) to Table 3 (which uses only their replies to the first two prompts) shows little qualitative difference in the findings.[6]

3.3 Robustness to Alternative Specifications

The author uses an ordinary least squares estimator on a categorical variable. While it is generally understood that using OLS will give similar results to alternative estimators, we start by estimating an multinomial logistic regression as the dependent variable – candidate rating – is a five-category ordinal variable.[7] We instead use a multinomial logistic regression and plot the predicted probabilities of respondents' support for a candidate at different levels of candidate competency for each level of the 'support' variable. The pooled model is visualised in Figure 6. Each panel corresponds to a different level of support for the candidate (from 'Very Unlikely' to 'Very Likely'), the x-axis plots the level of competence of the candidate, and the y-axis is the predicted probability of support. This plot tells a similar story to Figure 1, namely, that as candidate competence increases, the predicted probability of not supporting that candidate decreases (from 0.3 to 0.2 for undemocratic candidates and from ~ 0.26 to 0.16 for democratic candidates in the 'Very Unlikely' category, respectively), and the probability of supporting the candidate increases (from ~ 0.05 to ~ 0.15 for democratic candidates and from ~ 0.05 to ~ 0.12 for undemocratic candidates, respectively). In short, support increases for more competent candidates, even when they engage in 'undemocratic' behaviours.

[5] In the survey, respondents are asked a question that tests whether they have been paying attention to the study. Those that do not answer the question are considered to be inattentive.

[6] There is a difference in the level of significance, but this is understandable since one set of results uses only (approximately) 20% of the data as the other.

[7] It would seem natural to use logistic regression, however this model is inappropriate as the data fails the Brant test. This means that the parallel regression assumption—the assumption of correlation between dependent and independent variable not changing across the categories of the dependent variable—fails.

3.4 Robustness to Alternative Estimators

Conjoint Experiments allow for *causal inference* in many cases. One popular estimator for analyzing conjoint experiments is the average marginal component effect (AMCE) (Hainmueller, Hopkins, and Yamamoto, 2014; de la Cuesta, Egami, and Imai, 2022). This quantity can be thought of as the effect of a specific attribute, call it l, (e.g. adherence to Democratic values) on a choice or rating (e.g. likelihood of voting for a hypothetical candidate). Formally, the AMCE 'represents the marginal effect of attribute l averaged over the joint distribution of the remaining attributes.' (Hainmueller, Hopkins, and Yamamoto, 2014, pg. 10). An interesting feature of the AMCE is that it is 'defined as a function of the distribution of treatment components...' (Hainmueller, Hopkins, and Yamamoto, 2014, pg.11). By far the most common distribution of attributes in the analysis of conjoint experiments is the uniform distribution (see de la Cuesta, Egami, and Imai (2022, Figure 1). However, the researcher is free to chose other reasons (Hainmueller, Hall, and Snyder, 2015; Bansak et al., 2021). For example, assigning treatment attributes to match a (marginal) population distribution may improve external validity (Hainmueller, Hopkins, and Yamamoto, 2014; de la Cuesta, Egami, and Imai, 2022). Doing so changes slightly the calculation of the AMCE for an attribute (see Hainmueller, Hopkins, and Yamamoto (2014, Eqn. 4). Indeed, in Frederiksen (2022a), 'gender, age, and profession [attributes were assigned] using real-world, country-specific distributions of current and former ... incumbents to enhance external validity.' (pg. 1149). In the replication materials, the empirical (marginal) distribution is used in calculating AMCE. However, the research is free to use a different 'target distribution.' In this section we replicate estimated AMCE for attributes used in Frederiksen (2022b, Appendix D).[8] estimates AMCE of attributes using the issue-country-specific distributions given in Tables A1-A3 of Frederiksen (2022c), and the empirical distributions of policy distance between hypothetical candidate and survey respondent, co-partisanship (given in Table 5).

Hence, Figure 7 is a partial replication of Figure D6 in Frederiksen, 2022b with three notable exceptions. First, these estimates take into account the randomization distribution of variables used in the conjoint experiment. Second, all variables are treated as factors (e.g. `distance` takes one of seven levels). This is because pAMCE is only defined for target distributions over factor attributes (see Definition 3 of de la Cuesta, Egami, and Imai (2022)).[9] Lastly, Figure 7 does not contain a 'pooled' model, as does Figure D6 of Frederiksen (2022b). This is because the levels of attributes (e.g. age in which the values are a range in some countries and are specific values in others) are county-specific (see Table A1 of Frederiksen (2022c)). Hence, it is not clear what the target distribution in the pooled model should be.

Qualitatively, however, there is very little substantive difference between the AMCE and pAMCE of variables.[10] Comparing Figure 7 to Figure D6 yields the same qualitative conclusions (e.g. the effect of distance is negative and increasing in distance).[11] The magnitude of estimated effects are quite different, but this is likely explained by the variables being on different scales in the two figures. Hence, this demonstrates a robustness to (admittedly small) changes in the estimation of effects for different target distributions.

4 Discussion and Opportunities

We now turn to a brief discussion of our findings, limitations of our own replication, and the broader implications of this analysis. Overall, we find that Frederiksen, 2022a's findings are replicable and remarkably robust to a battery of robustness tests. Though only the pre-processed data are provided, tables and exhibits can be reproduced and replicated relatively easily. Of course, we would have benefited from having access to the raw data, but understand this this may be infeasible given the size of the raw data, or privacy constraints.

We highlight a number of potential avenues for future research, specifically using Frederiksen, 2022a's

[8] Specifically, non-democratic behavior, competency, policy distance between hypothetical candidate and survey respondent, co-partisanship, age, and gender). However, we estimate the population average marginal conditional effect (pAMCE) proposed in de la Cuesta, Egami, and Imai (2022) using the theoretical – rather than empirical – target distributions Specifically, Figure 7

[9] Though, one could easily adapt the definitions of AMCE and pAMCE to allow for more general distributions by replacing probability function, $Pr[t_{ijk,-l}, t_{i,-j,k}]$ to a density function and replacing the summation over profiles to an integral.

[10] It is not terribly surprising that controlling for background factors randomized independently of each other does not change the findings. We thank the author for pointing this out.

[11] Though, note the Democratic/Undemocratic variable is coded differently in the two analyses, hence the opposite sign of the estimated effect.

dataset. Given the data's richness—there are multiple tasks over measured over time, there is country variation, numerous outcomes measures, and a massive number of observations—there are numerous additional research questions that could be asked. For example, one could look at high-dimensional interactions of candidate attributes.

Lastly, our replication study has a number of limitations and areas for extension. Firstly, we have not replicated *all* of Frederiksen, 2022a's exhibits. Given there are two appendices, one of which is over 70 pages long, we decided that in the interest of space and time to replicate the main exhibits and only a handful of additional exhibits. Given that the exhibits we did replicate in our manuscript replicated perfectly, we feel confident that the remaining exhibits would be similarly replicable.

5 Conclusion

In this paper, we have reproduced, replicated, and assessed the robustness of Frederiksen, 2022a. We find that the author's analysis reproduces without much difficulty, and we are able to replicate their results from scratch in a different statistical software (namely R: rather than Stata). Moreover, we find that their primary results are very robust. Not only to different specifications, including multinomial logistic regression, but also to different sub-setting of data as well as different estimators. We commend the author for providing such a clear replication package. This paper contributes to the growing literature on the replication and robustness testing of existing studies in political science and economics, and provides a framework for additional exploratory analysis and robustness tests that may be beneficial for conjoint experiments (Gong and Hammar, 2023; Engel, Huber, and Nüß, 2022; Bonander, Strand, and Jakobsson, 2023). Future research could focus on assessing the conceptual replicability of this study by conducting this analysis on data from another country and meta-analysing studies that look at public attitudes towards candidate competence and democratic behaviour.

References

[1] Kirk Bansak et al. "Conjoint Survey Experiments". In: *Advances in Experimental Political Science*. Ed. by James Druckman and Donald P. Green. 1st ed. Cambridge University Press, Apr. 1, 2021, pp. 19–41.

[2] Denise Baron, Benjamin Lauderdale, and Jennifer Sheehy-Skeffington. "A Leader Who Sees the World as I Do: Voters Prefer Candidates Whose Statements Reveal Matching Social-Psychological Attitudes". In: *Political Psychology* 0.0 (2023).

[3] Pazit Ben-Nun. "Respondent Fatigue". In: *Encyclopedia of Survey Research Methods*. 2011, pp. 428–430.

[4] Christoph Beuthner et al. "Examining survey response styles in cross-cultural marketing research: A comparison between mexican and south korean respondents". In: *International Journal of Market Research* 60.3 (2018), pp. 257–267.

[5] Carl Bonander, Gabriella Chauca Strand, and Niklas Jakobsson. "Direct replication and additional sensitivity analyses for Altindag et al. (2022): A replication report from the Oslo Replication Games". 2023.

[6] Rosie Campbell and Philip Cowley. "Rich man, poor man, politician man: Wealth effects in a candidate biography survey experiment". In: *The British Journal of Politics and International Relations* 16.1 (2014), pp. 56–74.

[7] Rosie Campbell and Philip Cowley. "The impact of parental status on the visibility and evaluations of politicians". In: *The British Journal of Politics and International Relations* 20.3 (2018), pp. 753–769.

[8] Rosie Campbell and Philip Cowley. "What voters want: Reactions to candidate characteristics in a survey experiment". In: *Political Studies* 62.4 (2014), pp. 745–765.

[9] Nicholas Carnes and Noam Lupu. "Do voters dislike working-class candidates? Voter biases and the descriptive underrepresentation of the working class". In: *American Political Science Review* 110.4 (2016), pp. 832–844.

[10] P Cowley et al. "Legislator dissent as a valence signal". In: *British Journal of Political Science* (2016).

[11] Brandon de la Cuesta, Naoki Egami, and Kosuke Imai. "Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution". In: *Political Analysis* 30.1 (Jan. 2022), pp. 19–45.

[12] James N. Druckman and Donald P. Green. *Advances in Experimental Political Science*. Cambridge University Press, 2021.

[13] Andrew C Eggers, Nick Vivyan, and Markus Wagner. "Corruption, accountability, and gender: Do female politicians face higher standards in public life?" In: *The Journal of Politics* 80.1 (2018), pp. 321–326.

[14] Julia F Engel, Christoph Huber, and Patrick Nüß. "Replication Report: How Do Beliefs About the Gender Wage Gap Affect the Demand for Public Policy?" 2022.

[15] Kristian Vrede Skaaning Frederiksen. "Does Competence Make Citizens Tolerate Undemocratic Behavior?" In: *American Political Science Review* 116.3 (2022), pp. 1147–1153.

[16] Kristian Vrede Skaaning Frederiksen. "Does Competence Make Citizens Tolerate Undemocratic Behavior? Dataverse Appendix". In: *American Political Science Review* 116.3 (2022), pp. 1147–1153.

[17] Kristian Vrede Skaaning Frederiksen. "Does Competence Make Citizens Tolerate Undemocratic Behavior? Supplementary Materials". In: *American Political Science Review* 116.3 (2022), pp. 1147–1153.

[18] Da Gong and Olle Hammar. "Replication Report: A Comment on Gethin, Martínez- Toledano Piketty (2022)". 2023.

[19] Eric Guntermann and Gabriel S Lenz. "Replication of "Re-Assessing Elite-Public Gaps in Political Behavior" by Joshua Kertzer". 2022.

[20] Jens Hainmueller, Andrew B. Hall, and James M. Snyder. "Assessing the External Validity of Election RD Estimates: An Investigation of the Incumbency Advantage". In: *The Journal of Politics* 77.3 (July 2015), pp. 707–720.

[21] Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments". In: *Political Analysis* 22.1 (Jan. 1, 2014), pp. 1–30.

[22] Sara B Hobolt and Toni Rodon. "Cross-cutting issues and electoral choice. EU issue voting in the aftermath of the Brexit referendum". In: *Journal of European Public Policy* 27.2 (2020), pp. 227–245.

[23] Anthony Kevins. "Race, class, or both? Responses to candidate characteristics in Canada, the UK, and the US". In: *Politics, Groups, and Identities* 9.4 (2021), pp. 699–720.

[24] Thomas J. Leeper, Sara B. Hobolt, and James Tilley. "Measuring Subgroup Preferences in Conjoint Experiments". In: *Political Analysis* 28.2 (2020), pp. 207–221.

[25] Gabriele Magni and Andrew Reynolds. "The Persistence of Prejudice: Voters Strongly Penalize Candidates with HIV". In: *Political Behavior* (2022), pp. 1–20.

[26] Gabriele Magni and Andrew Reynolds. "Voter preferences and the political underrepresentation of minority groups: Lesbian, gay, and transgender candidates in advanced democracies". In: *Journal of Politics* 83.4 (2021), pp. 1199–1215.

[27] Nicole S Martin and Scott Blinder. "Biases at the ballot box: How multiple forms of voter discrimination impede the descriptive and substantive representation of ethnic minority groups". In: *Political Behavior* 43.4 (2021), pp. 1487–1510.

[28] Stefanie Reher. "How do voters perceive disabled candidates?" In: *Frontiers in Political Science* 2 (2021), p. 634432.

[29] Sparsha Saha and Ana Catalano Weeks. "Ambitious women: Gender and voter perceptions of candidate ambition". In: *Political Behavior* 44.2 (2022), pp. 779–805.

[30] Nick Vivyan and Markus Wagner. "House or home? Constituent preferences over legislator effort allocation". In: *European Journal of Political Research* 55.1 (2016), pp. 81–99.

[31] Nick Vivyan and Markus Wagner. "What do voters want from their local MP?" In: *The Political Quarterly* 86.1 (2015), pp. 33–40.

[32] Nick Vivyan, Markus Wagner, et al. "Do humble beginnings help? How politician class roots shape voter evaluations". In: *Electoral Studies* 63 (2020), p. 102093.

Figure 1: Replication of Figure 1, Frederiksen, 2022a. The x-axis is a five-point measure of candidate competency, captured by the vignettes in Table 1 of the paper. The y-axis is the marginal mean value of a five-point measure of respondents' level of support for each candidate. The line colours discriminate between democratic (black) and undemocratic (grey) candidates. The first panel is a pooled model (that is, including all country-year observations), and each subsequent panels are results disaggregated by country. Lower panels present whether the effects of undemocratic behavior for incompetent and competent candidates differ from those among average competence (a response of three, [category omitted]).

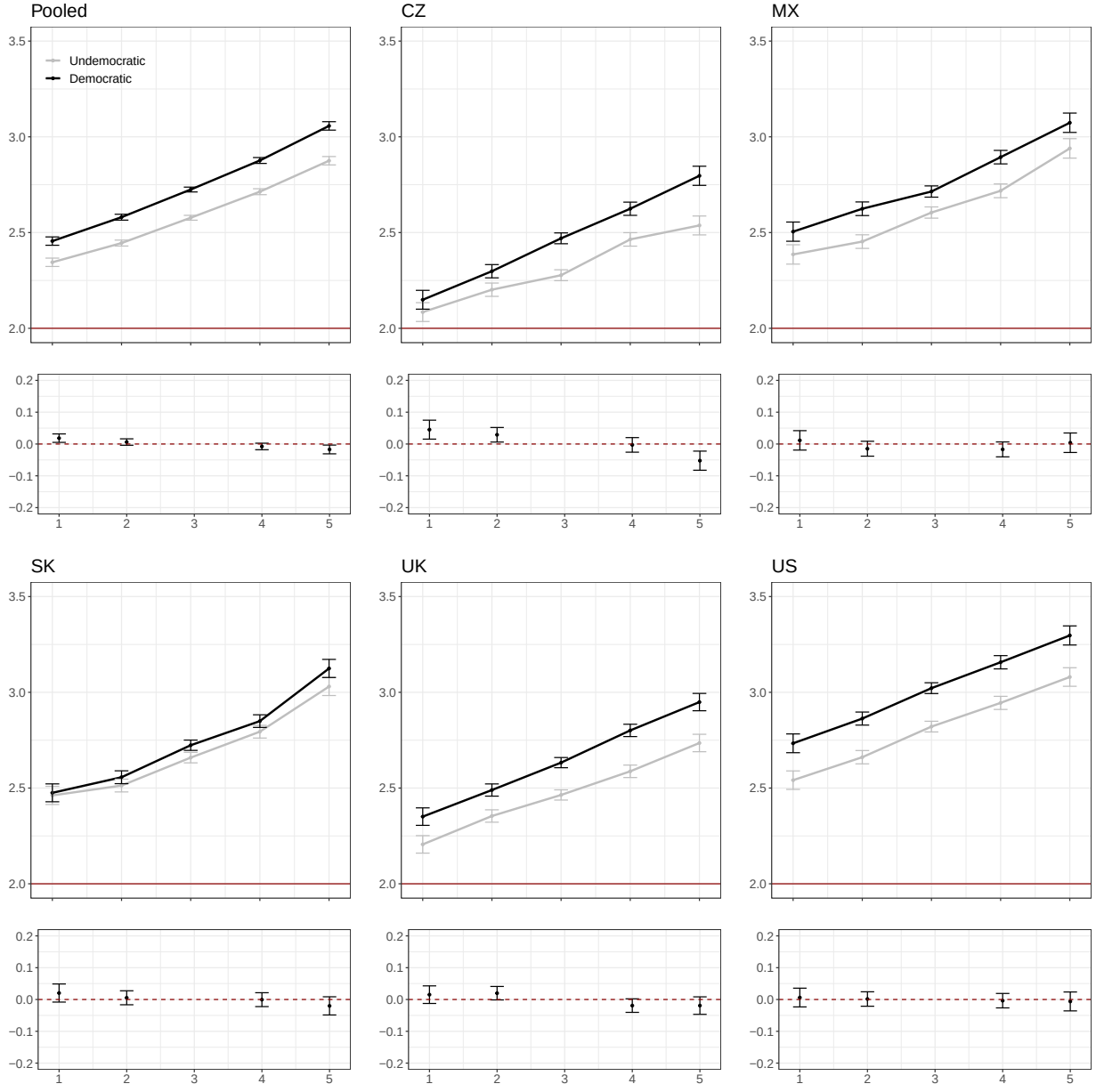

Table 1: (Replication of Table A1 (Frederiksen, 2022b)) Average effects of undemocratic behavior and competence in the Czech Republic, Mexico, South Korea, the United Kingdom, and the United States. Candidate support is the dependent variable in all models.

	Pooled	CZ	MX	SK	UK	US
Undemocratic behavior	-0.15^{***}	-0.16^{***}	-0.14^{***}	-0.06^{***}	-0.17^{***}	-0.20^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Very incompetent	-0.25^{***}	-0.26^{***}	-0.21^{***}	-0.22^{***}	-0.27^{***}	-0.28^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Incompetent	-0.14^{***}	-0.12^{***}	-0.12^{***}	-0.16^{***}	-0.13^{***}	-0.16^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Competent	0.14^{***}	0.17^{***}	0.15^{***}	0.13^{***}	0.15^{***}	0.13^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Very Competent	0.31^{***}	0.29^{***}	0.35^{***}	0.39^{***}	0.29^{***}	0.27^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	2.72^{***}	2.45^{***}	2.73^{***}	2.72^{***}	2.64^{***}	3.02^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
R^2	0.02	0.02	0.02	0.02	0.02	0.02
Adj. R^2	0.02	0.02	0.02	0.02	0.02	0.02
Num. obs.	267795	47221	55167	50002	55299	60106
RMSE	1.36	1.29	1.43	1.26	1.29	1.44
N Clusters	14058	2481	2845	2691	2882	3159

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: (Replication of Table A2 (Frederiksen, 2022b)) Effects of undemocratic behavior interacted by, candidate competence in the Czech Republic, Mexico, South Korea, the United Kingdom, and the United States. Candidate support is the dependent variable in all models.

	Pooled	CZ	MX	SK	UK	US
Undemocratic behavior	-0.15^{***}	-0.19^{***}	-0.11^{***}	-0.06^{***}	-0.17^{***}	-0.20^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Very incompetent	-0.27^{***}	-0.32^{***}	-0.21^{***}	-0.25^{***}	-0.28^{***}	-0.29^{***}
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Incompetent	-0.14^{***}	-0.17^{***}	-0.09^{***}	-0.17^{***}	-0.14^{***}	-0.16^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Very competent	0.15^{***}	0.15^{***}	0.18^{***}	0.13^{***}	0.17^{***}	0.14^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Undemocratic x Very incompetent	0.04^{*}	0.13^{**}	-0.01	0.05	0.02	0.01
	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Undemocratic x Incompetent	0.01	0.10^{**}	-0.06	0.02	0.03	-0.00
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Undemocratic x Competent	-0.02	0.03	-0.07^{*}	0.01	-0.04	-0.01
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Undemocratic x Very competent	-0.03	-0.07	-0.02	-0.03	-0.04	-0.02
	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
R^2	0.02	0.02	0.02	0.02	0.02	0.02
Adj. R^2	0.02	0.02	0.02	0.02	0.02	0.02
Num. obs.	267795	47221	55167	50002	55299	60106
RMSE	1.36	1.29	1.43	1.26	1.29	1.44
N Clusters	14058	2481	2845	2691	2882	3159

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Table 2 using only first two responses, per respondent. Effects of undemocratic behavior interacted by candidate competence in the Czech Republic, Mexico, South Korea, the United Kingdom, and the United States. Candidate support is the dependent variable in all models. Note: only first two (out of 10) prompts are used in these calculations

	Pooled	CZ	MX	SK	UK	US
Undemocratic behavior	-0.14^{***}	-0.18^{***}	-0.13^{**}	-0.01	-0.14^{**}	-0.23^{***}
	(0.02)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)
Very incompetent	-0.32^{***}	-0.40^{***}	-0.30^{***}	-0.34^{***}	-0.35^{***}	-0.25^{***}
	(0.03)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)
Incompetent	-0.19^{***}	-0.22^{***}	-0.13^{*}	-0.23^{***}	-0.13^{**}	-0.22^{***}
	(0.02)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Very competent	0.21^{***}	0.18^{**}	0.18^{**}	0.19^{***}	0.28^{***}	0.23^{***}
	(0.02)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Undemocratic x Very incompetent	0.04	0.16	0.04	0.09	0.07	-0.12
	(0.04)	(0.09)	(0.10)	(0.09)	(0.09)	(0.09)
Undemocratic x Incompetent	0.02	0.12	0.04	0.01	-0.07	0.04
	(0.03)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
Undemocratic x Competent	-0.03	0.03	-0.05	-0.05	-0.10	-0.01
	(0.03)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
Undemocratic x Very competent	-0.03	-0.19	0.04	-0.07	-0.04	0.02
	(0.04)	(0.10)	(0.10)	(0.09)	(0.09)	(0.10)
R^2	0.02	0.02	0.02	0.03	0.03	0.02
Adj. R^2	0.02	0.02	0.02	0.02	0.03	0.02
Num. obs.	53940	9495	11082	10177	11100	12086
RMSE	1.42	1.37	1.49	1.32	1.38	1.48
N Clusters	13887	2456	2828	2621	2856	3126

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Sample distribution of attributes: US, UK, and CZ. Age is drawn randomly from probability-specified intervals. This Table shows the empirical analog of Tables A1, A2, and A3 in Frederiksen, 2022b.

	Attribute Category	CZ	UK	US	MX	SK
1	43-57	0.333				
2	58-67	0.334				
3	68-77	0.334				
4	44-53		0.332			
5	54-57		0.336			
6	58-61		0.332			
7	40-49			0.198		
8	50-59			0.2		
9	60-62			0.199		
10	63-66			0.201		
11	67-75			0.201		
12	Female	0.083	0.286	0.202	0.126	0.055
13	Male	0.917	0.714	0.798	0.874	0.945
14	Actor/actress	0.169				
15	Journalist	0.252	0.224		0.032	0.055
16	Political Career	0.084	0.222	0.101		0.112
17	Professor	0.496				0.055
18	Academic				0.031	
19	Accountant				0.124	
20	Business Administration				0.061	
21	Civil Servant		0.111	0.101	0.094	0.056
22	Engineer				0.127	0.056
23	Lawyer		0.221	0.301	0.407	0.387
24	Professional Sports				0.031	
25	Self-employed			0.098	0.093	0.167
26	Army General					0.056
27	Company Director					0.055
28	Banker		0.222			
29	Company Founder/Director			0.399		
30	ANO 2011	0.336				
31	ODS (Občanská demokratická strana)	0.333				
32	ČSSD (Česká strana sociálně demokratická)	0.331				
33	MORENA				0.249	
34	PAN				0.249	
35	PRD				0.251	
36	PRI				0.251	
37	Democratic Party of Korea					0.499
38	United Future Party					0.501
39	Conservatives		0.502			
40	Labour		0.498			
41	Democrat			0.501		
42	Republican			0.499		
43	Decrease income tax on 10 percent richest	0.167	0.163	0.166	0.166	0.169
44	Decrease power of labor unions	0.169		0.169		0.167
45	Decrease public welfare spending	0.164	0.165	0.166	0.165	0.167
46	Increase income tax on 10 percent richest	0.163	0.168	0.165	0.169	0.167
47	Increase power of labor unions	0.168		0.165		0.165
48	Increase public welfare spending	0.169	0.169	0.168	0.158	0.166

49	Prevent universal access to public colleges				0.166	
50	Provide universal access to public colleges				0.177	
51	Decrease power of trade unions		0.168			
52	Increase power of trade unions		0.166			
53	Allow illegal immigrants to apply for citizenship	0.166	0.167	0.168		
54	Increase efforts to arrest and eventually deport illegal immigrants	0.167	0.165	0.164		
55	Make it easier for people of the same sex to marry each other	0.166	0.168	0.164		
56	Make it easier for women to get an abortion	0.166	0.166	0.168		
57	Make it harder for people of the same sex to marry each other	0.168	0.166	0.168		
58	Make it harder for women to get an abortion	0.168	0.168	0.168		
59	Legalize same-sex marriage nationally				0.169	0.167
60	Make abortion law more strict				0.166	0.167
61	Prohibit same-sex marriage nationally				0.167	0.168
62	Provide amnesty to low-level drug offenders				0.166	
63	Punish all drug-related crime harsher				0.165	
64	Relax abortion law				0.166	0.168
65	Decrease funds to the army					0.166
66	Increase funds to the army					0.165
67	Said court rulings by judges appointed by opposing parties should be adhered to	0.125	0.127	0.124	0.125	0.126
68	Said court rulings by judges appointed by opposing parties should be ignored	0.125	0.124	0.125	0.125	0.123
69	Said it is acceptable to harass journalists that do not reveal sources	0.125	0.123	0.124	0.124	0.124
70	Said it is legitimate to fight political opponents in the streets if one feels provoked	0.126	0.123	0.127	0.125	0.125
71	Said it is unacceptable to fight political opponents in the streets even though one feels provoked	0.123	0.126	0.125	0.127	0.127
72	Said it is unacceptable to harass journalists even though they do not reveal sources	0.125	0.126	0.123	0.123	0.124
73	Supported a proposal to preserve existing polling-stations in all areas	0.128	0.124	0.125	0.126	0.126
74	Supported a proposal to reduce polling stations in areas that support opposing parties	0.122	0.127	0.127	0.125	0.125
75	Bad at handling economic matters	0.336	0.338	0.335	0.334	0.335
76	Good at handling economic matters	0.33	0.331	0.331	0.333	0.335
77	Neither good nor bad reputation on economic matters	0.334	0.331	0.334	0.333	0.33
78	Bad at fighting corruption	0.337	0.331	0.333	0.335	0.334
79	Good at fighting corruption	0.334	0.336	0.332	0.331	0.333
80	Neither good nor bad reputation on fighting corruption	0.329	0.333	0.335	0.335	0.333

Table 5: The sample distribution of certain attributes used in Appendix C and D of Frederiksen (2022b)

	Attribute Category	CZ	MX	SK	UK	US
1	2	0.174	0.126	0.14	0.162	0.135
2	3	0.22	0.125	0.254	0.224	0.208
3	4	0.242	0.308	0.233	0.248	0.265
4	5	0.115	0.099	0.108	0.112	0.1
5	6	0.065	0.061	0.064	0.056	0.063
6	Completely Aligned Positions	0.142	0.181	0.16	0.148	0.161
7	Completely Diverting Positions	0.042	0.1	0.04	0.051	0.069
8	1	0.142	0.181	0.16	0.148	0.161
9	1.67	0.174	0.126	0.14	0.162	0.135
10	2.33	0.22	0.125	0.254	0.224	0.208
11	3	0.242	0.308	0.233	0.248	0.265
12	3.67	0.115	0.099	0.108	0.112	0.1
13	4.33	0.065	0.061	0.064	0.056	0.063
14	5	0.042	0.1	0.04	0.051	0.069
15	Dislike candidate's party a great deal	0.335	0.32	0.226	0.224	0.23
16	Dislike candidate's party somewhat	0.175	0.163	0.123	0.156	0.123
17	Like candidate's party a great deal	0.08	0.103	0.104	0.131	0.277
18	Like candidate's party somewhat	0.168	0.161	0.201	0.264	0.214
19	Neither like nor dislike candidate's party	0.243	0.253	0.345	0.225	0.156
20	43	0.022				0.02
21	44	0.022	0.063		0.033	0.019
22	45	0.022	0.032		0.033	0.02
23	46	0.022	0.032		0.033	0.021
24	47	0.022	0.031		0.033	0.02
25	48	0.022	0.031		0.033	0.02
26	49	0.022	0.063	0.056	0.034	0.02
27	50	0.023	0.094		0.034	0.02
28	51	0.022	0.031		0.034	0.02
29	52	0.023	0.095		0.032	0.021
30	53	0.022	0.031		0.033	0.02
31	54	0.022	0.031	0.055	0.084	0.02
32	55	0.022	0.031	0.112	0.083	0.02
33	56	0.023	0.063	0.056	0.084	0.02
34	57	0.022	0.094		0.086	0.019
35	58	0.034	0.03		0.083	0.02
36	59	0.034		0.055	0.083	0.02
37	60	0.034	0.032	0.111	0.082	0.065
38	61	0.033	0.062		0.084	0.067
39	62	0.033	0.031	0.054		0.067
40	63	0.032		0.057		0.051
41	64	0.033		0.056		0.05
42	65	0.034		0.055		0.051
43	66	0.033		0.055		0.05
44	67	0.034		0.055		0.022
45	68	0.033		0.056		0.022
46	69	0.033	0.061			0.021
47	70	0.034				0.023
48	71	0.033				0.023
49	72	0.034		0.055		0.023

50	73	0.034		0.056		0.023
51	74	0.034				0.023
52	75	0.033				0.022
53	76	0.033				
54	77	0.034		0.056		
55	39		0.031			
56	40		0.033			0.02
57	41					0.02
58	42					0.019
59	4.3	0.022				0.02
60	4.4	0.022	0.063		0.033	0.019
61	4.5	0.022	0.032		0.033	0.02
62	4.6	0.022	0.032		0.033	0.021
63	4.7	0.022	0.031		0.033	0.02
64	4.8	0.022	0.031		0.033	0.02
65	4.9	0.022	0.063	0.056	0.034	0.02
66	5	0.023	0.094		0.034	0.02
67	5.1	0.022	0.031		0.034	0.02
68	5.2	0.023	0.095		0.032	0.021
69	5.3	0.022	0.031		0.033	0.02
70	5.4	0.022	0.031	0.055	0.084	0.02
71	5.5	0.022	0.031	0.112	0.083	0.02
72	5.6	0.023	0.063	0.056	0.084	0.02
73	5.7	0.022	0.094		0.086	0.019
74	5.8	0.034	0.03		0.083	0.02
75	5.9	0.034		0.055	0.083	0.02
76	6	0.034	0.032	0.111	0.082	0.065
77	6.1	0.033	0.062		0.084	0.067
78	6.2	0.033	0.031	0.054		0.067
79	6.3	0.032		0.057		0.051
80	6.4	0.033		0.056		0.05
81	6.5	0.034		0.055		0.051
82	6.6	0.033		0.055		0.05
83	6.7	0.034		0.055		0.022
84	6.8	0.033		0.056		0.022
85	6.9	0.033	0.061			0.021
86	7	0.034				0.023
87	7.1	0.033				0.023
88	7.2	0.034		0.055		0.023
89	7.3	0.034		0.056		0.023
90	7.4	0.034				0.023
91	7.5	0.033				0.022
92	7.6	0.033				
93	7.7	0.034		0.056		
94	3.9		0.031			
95	4		0.033			0.02
96	4.1					0.02
97	4.2					0.019

Figure 2: Histogram of seconds spent completing the survey for each respondent, where each panel is a different country. The x-axis captures the time taken in seconds, while the y-axis is the frequency of observations.

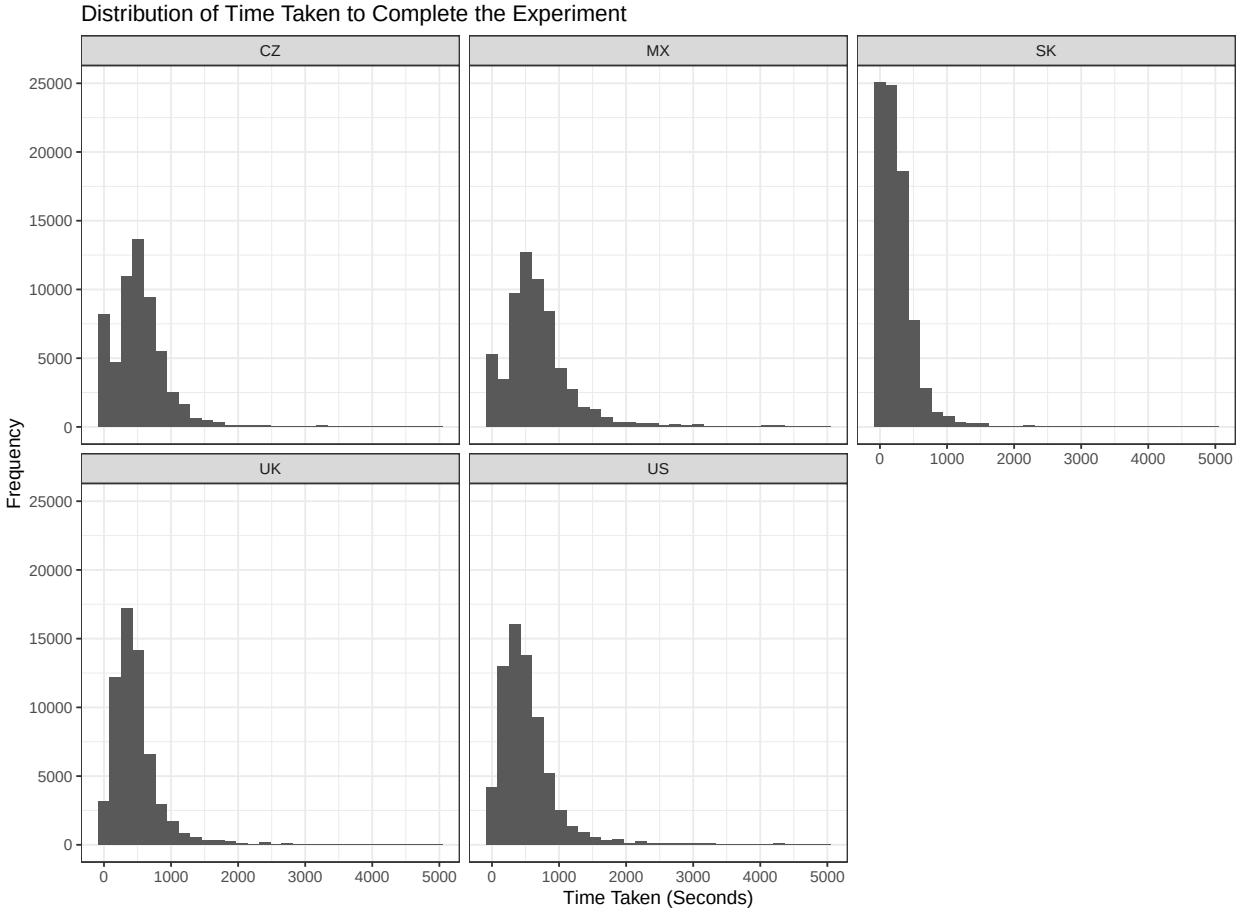

Figure 3: This figure shows the mean difference in candidate pair choices across the ten different tasks the respondents are asked to complete. The x-axis is the task number, and the y-axis is the mean difference in rating between the two candidates generated for each task. A more detailed description of the logic behind this plot can be found in section 3.1.

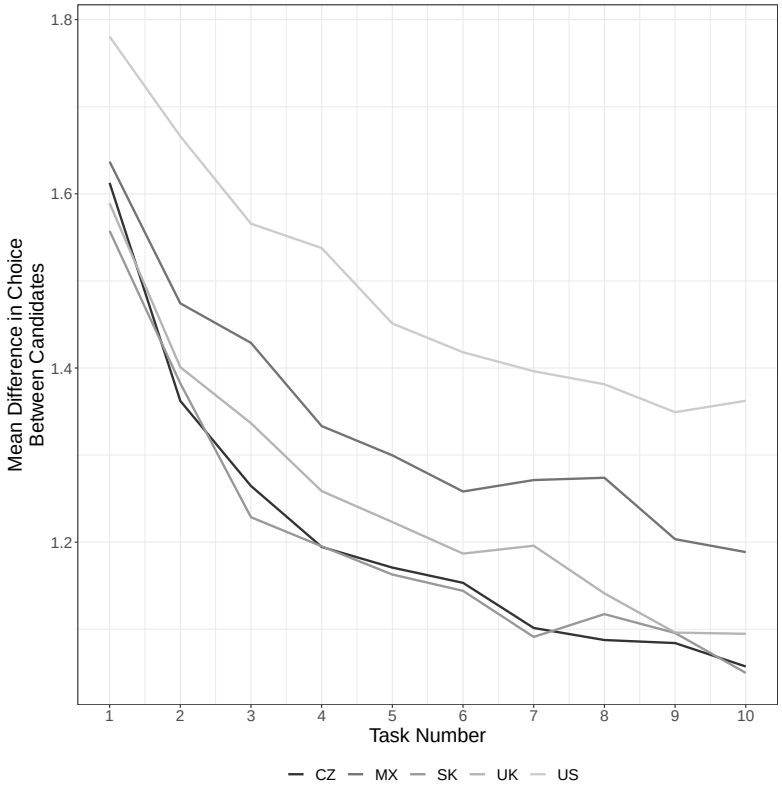

Figure 4: Robustness to Respondent Time Taken to Complete the Survey. This figure plots the point estimates and confidence intervals from the regression specified in equation 1, where each coloured point represents a different truncation of the data (starting with the removal of the upper and lower ten percentiles of response time). The x-axis is the point estimate, and the y-axis represents each coefficient from the regression.

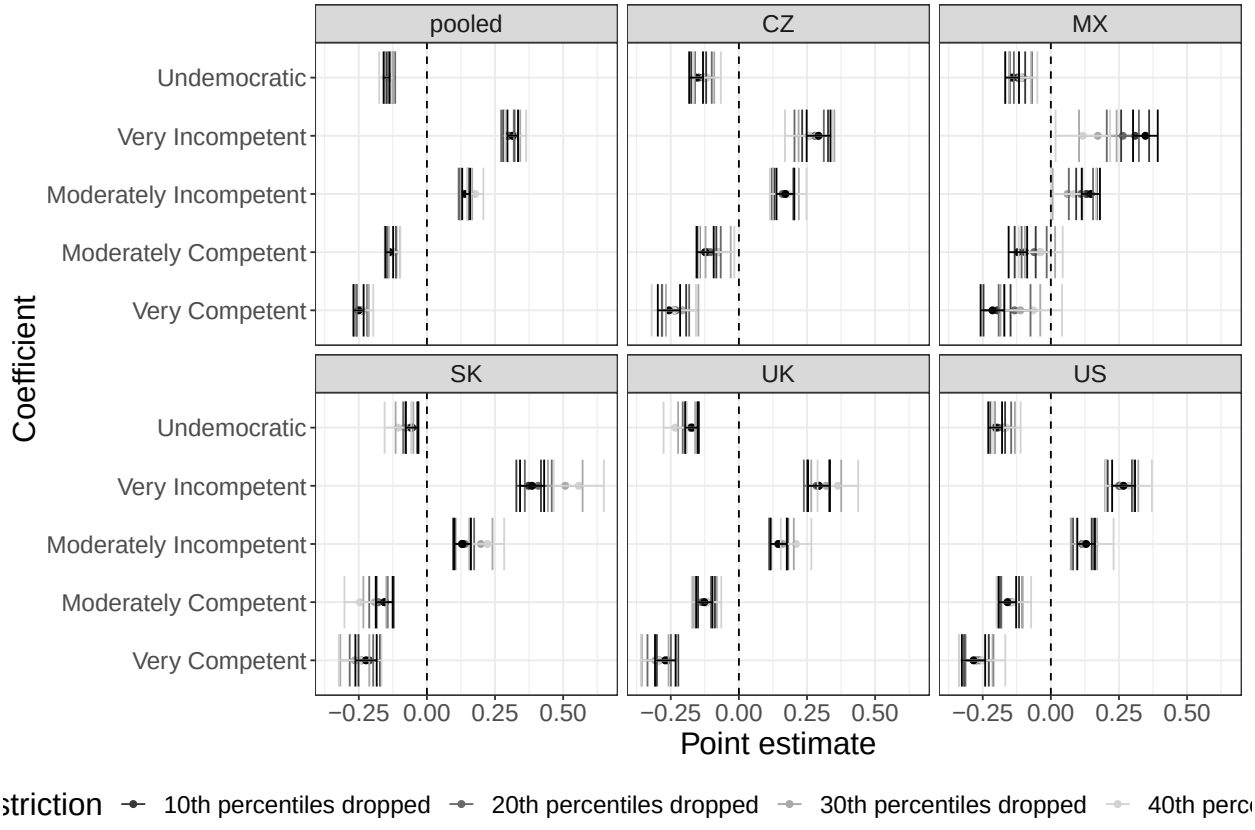

Figure 5: Robustness to Removal of 'Inattentive' Respondents. This figure presents point estimates from the regression specified in equation 1, except we subset the data to not include 'inattentive' respondents. The x-axis is the point estimate, and the y-axis represents each coefficient from the regression.

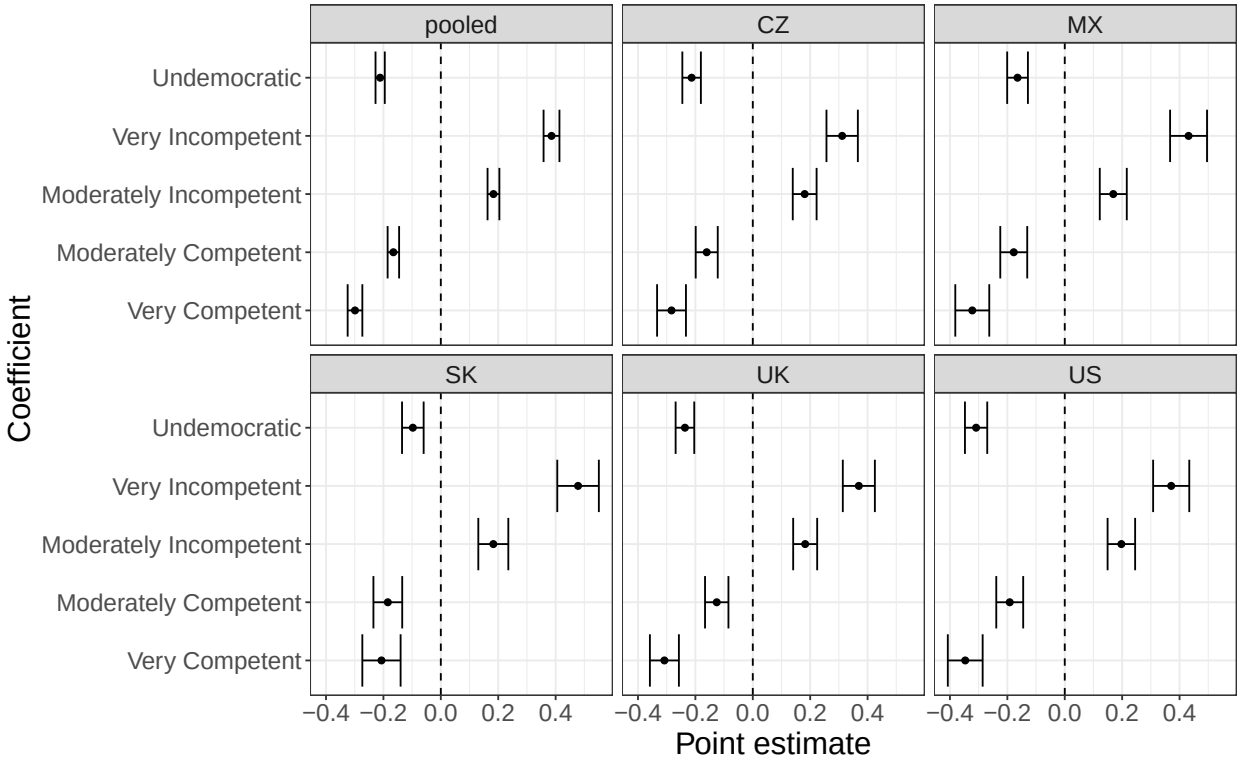

Figure 6: Predicted Probabilities of Support for Candidates (Pooled Model). This figure presents the predicted probabilities of candidate support obtained from a multinomial logistic regression for each level of candidate support. The x-axis is a five-point measure of candidate competence. The y-axis is the predicted probability of the respondents' support for the candidate. The line colour discriminates between democratic and undemocratic candidates, while each facet captures the each category of support (from 'Very Unlikely' to support to 'Very Likely' to support).

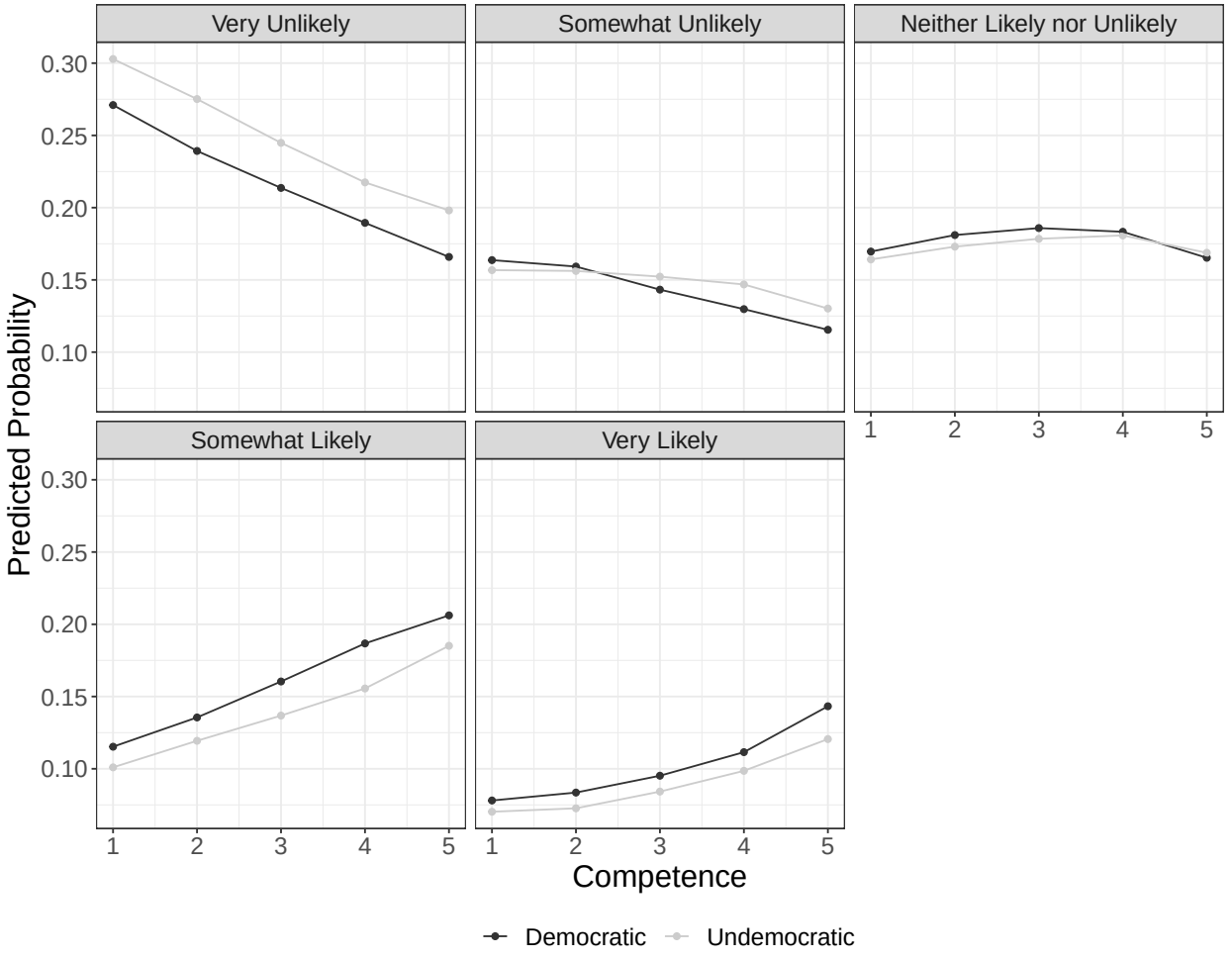

Figure 7: Partial replication of Figure D6 (Frederiksen, 2022b) using the population AMCE estimator of de la Cuesta, Egami, and Imai (2022)