INSTITUTE for REPLICATION

No. 87
I4R DISCUSSION PAPER SERIES

Temporary Disenfranchisement Revisited: A Report from the 2023 Montréal Replication Games on the Robustness of Recent Findings in the APSR

Joris Frese

Alexandros Christos Gkotinakos

Pau Grau

Matthew Hepplewhite

November 2023

INSTITUTE for REPLICATION

I4R Discussion Paper Series

I4R DP No. 87

Temporary Disenfranchisement Revisited: A Report from the 2023 Montréal Replication Games on the Robustness of Recent Findings in the APSR

Joris Frese[1], Alexandros Christos Gkotinakos[2], Pau Grau[1], Matthew Hepplewhite[3]

[1]*European University Institute, Florence/Italy*
[2]*Aristotle University of Thessaloniki/Greece*
[3]*University of Oxford/Great Britain*

November 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

Temporary Disenfranchisement Revisited: A Report from the 2023 Montréal Replication Games on the Robustness of Recent Findings in the APSR

Joris Frese[†]
joris.frese@eui.eu

Alexandros Christos Gkotinakos[§]
alexandgm@gapps.auth.gr

Pau Grau[†]
pau.grau@eui.eu

Matthew Hepplewhite[¶] [*]
matthew.hepplewhite@politics.ox.ac.uk

09 November 2023

† European University Institute

§ Aristotle University of Thessaloniki

¶ University of Oxford

[*]The author list is alphabetically ordered. All authors have contributed equally.

Abstract

Leininger et al. (2023) study the political consequences of temporary disenfranchisement. Taking advantage of differentiated voting elegibility thresholds applying in different elections in Germany, they analyze how first-time voters react when losing eligibility in a follow-up election. They exploit this setting in a difference-in-differences design using panel data. They find that temporary disenfranchisement decreases perceived external efficacy by 0.19 points on a five-point Likert scale and satisfaction with democracy by 0.14 points. Both results are statistically significant at the five-percent level. In contrast, internal efficacy and political interest remain unaffected by the treatment, and regaining voting eligibility is not associated with statistically significant changes in respondents' attitudes.

This report focuses on the computational reproducibility and robustness replicability of these findings. To assess the paper's reproducibility, we first attempt to reproduce the paper's estimates and figures using the author's replication materials. In a second step, we perform several robustness checks by means of alternative difference-in-differences specifications using coarsened exact matching and entropy balancing, and a closer examination of panel attrition. Overall, we find complete reproducibility of the original replication materials. Our robustness checks confirm the sign congruence and significance of coefficients reported in the original paper. We raise the issue of potential bias due to differential panel attrition rates between treated and untreated respondents.

GitHub Repository: https://github.com/paugrau/2023_replication_montreal

Keywords: Replication, Matching, Attrition

1 Introduction

The following paper was produced at the Replication Games held on 14 June 2023 at the Université du Québec à Montréal (UQAM), 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 Montréal event was the eighth iteration of the Replication Games format, the previous seven Games having been held in Oslo, Calgary, Toronto, Nottingham, Vienna, Ottawa, and Melbourne (https://i4replication.org/games.html).

For this event, researchers are grouped based on their disciplinary affiliation and their research interests. Each group is given a selection of recent papers from the leading journals in their discipline, which, in our case, are the leading political science journals, *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics*. Each group then chooses one paper for which the original analyses are replicated and enriched with additional analyses. Our group decided to replicate the APSR paper "Temporary Disenfranchisement: Negative Side Effects of Lowering the Voting Age" by Leininger et al. (2023). A short primer on the theoretical framework of this paper is given in the second section of the report.

Following this theoretical discussion, we next conduct a computational reproduction of all the analyses documented in the replication files by running the original code on the original data. All replication files for Leininger et al. (2023) are freely available on https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/T5LYWS. We thank the authors for their transparent and public data sharing and confirm the full reproducibility of the original code, as our outputs are identical to those reported in the APSR.

Afterwards, we conduct a series of additional robustness checks and alternative model specifications to scrutinize the replicability of the results more carefully. First, we replicate the main difference-in-differences (DiD) analysis on balanced samples using coarsened exact matching and entropy balancing, leading us to practically identical conclusions as the unweighted original analysis and the Appendix analysis on a matched sample (nearest-neighbor matching). Next, we take a closer look at panel attrition. Here, we find significant differences in panel attrition between the treatment group and the control group. We give some further suggestive evidence that this differential panel attrition might bias the results due to differences between panel dropouts and remainers.

In sum, the results of Leininger et al. (2023) are robust to our replication analyses and alternative model specifications. The only concerning element of their research design is the differential panel attrition between treated and untreated respondents, which is however largely out of the control of the researchers. All of our replication analyses are included in the public GitHub repository linked under the Abstract.

2 Theory & Background

Before delving deeper into the details of our replication, it is worth familiarizing ourselves with the paper's motivation, theoretical background, case study, and contribution. Starting with the motivation, the authors identify a gap in the existing literature on young people's political behavior. Several studies have looked into patterns of formal political participation and young people's turnout propensity. Most relevant to this paper, the authors summarize, the literature has hinted towards relatively lower participation rates for young (compared to older) voters, but with a potential nuance of 18–19-year-old voters (compared to voters in their early 20s) being more enthusiastic about voting, and the positive effect of turnout experience on habituating (future) turnout and political engagement. Moreover, less is known about the opposite scenario, the effect of disenfranchisement, that is, on future political engagement. Drawing on the literature on democratic responsiveness and citizens' evaluation of the political process, the authors hypothesize that losing one's right to vote may negatively impact those evaluations. Prior voting experience coupled with downstream voting ineligibility could lead to feelings of resentment towards the political system.

While disenfranchisement results from diverse circumstances, the authors draw attention to the specific phenomenon of temporary disenfranchisement. In contrast to other cases where the right to vote gets restricted for criminal behavior or the like, temporary disenfranchisement has nothing to do with the individual voter per se in this case. It is purely an artifact of inconsistent voting eligibility laws and occurs when a differential age threshold applies to different types of elections.

The case of Germany, a federal state divided into 16 Länder, with elections at the national, supranational, and subnational levels, is a prime setting to study this phenomenon. To do that, the authors collected panel data through a three-wave survey in the state of Schleswig-Holstein, where, in 2017, three types of elections took place within a one-year period. In particular, following the state-level elections, the national elections followed just a few months after, tailed by the municipal elections. What's important, from a research design perspective, is that voting age laws are unharmonized between those elections. While for the state and municipal elections the voting age is 16, voting in the national elections is not possible until the age of 18. As a result, several young people who voted in the state elections of May 2017 were practically disenfranchised at the national level in the immediate September legislative election and re-enfranchised again in May 2018.

Leininger et al. (2023) relate this case to the discussion on the potential effects of disenfranchisement on political engagement to ask: what was the effect of temporary disenfranchisement on aspects of political engagement? At this point, the authors introduce the conceptual distinction between internal and external political efficacy to develop four expectations. They argue that attitudes relating to internal political efficacy are formed in the early socialization years and remain relatively stable over time. By contrast, evaluations of the political system can be subject to continuous belief updating. Thus, the authors posit that while temporary disenfranchisement can be expected to cause a decrease in satisfaction with democracy and external efficacy, it should not affect internal efficacy or political interest. Further, in considering the potential countervailing effect of regaining the right to vote, Leininger et al. (2023) draw on prospect theory to argue that, since losses tend to outweigh gains in utility, the net effect of temporary disenfranchisement, followed by re-enfranchisement, should be, on average, negative.

Through a difference-in-differences (DiD) research design, the authors empirically advance the above research questions and test their hypotheses by splitting the sample into those young voters who were eligible in all three elections in 2017 (control group) and those who were eligible in the state-level election of May 2017 but ineligible in the following national election (treatment group). Leininger et al. (2023) demonstrate how losing eligibility has a negative and statistically significant effect on external efficacy and satisfaction with democracy. This effect is estimated through the first and second wave of the panel as a 0.19 and 0.14 point decrease (on a five-point Likert scale), respectively. Further, at the third wave, respondents were enfranchised again for the municipal elections, making it possible to test the re-enfranchisement hypotheses. The effects of re-enfranchisement, while positive, do not pass the conventional statistical thresholds for significance. The net effect of this voting age inconsistency appears to result in an overall, net decrease in external political efficacy (though the net effect is not statistically significant either). Thus, the authors make a contribution to the literature by showing the negative effects of temporary disenfranchisement on external political efficacy. Their findings have implications for policy design of voting laws that are especially relevant for the subnational/local levels.

3 Computational Reproducibility

To start our replication exercise, we first investigate the computational reproducibility of the original code and data, as provided by the authors. As mentioned, Leininger et al. (2023) have made all code and data publicly available. In rerunning the code, we were able to successfully reproduce all data transformations, estimates, tables, and figures with no discrepancies or computational problems. The code is free of any errors. It is especially commendable that great care has been taken to ensure reproducibility by means of package installation facilitation and literate programming.

4 Robustness Replication

In the following subsection, we first present the results of robustness checks using state of the art matching and weighting schemes, including coarsened exact matching and entropy balancing. Afterward, we assess panel attrition rates and compare these between treated and untreated respondents. We also provide some examples of how differential attrition might bias treatment effects.

4.1 Coarsened Exact Matching & Entropy Balancing

In their Appendix, Leininger et al. (2023) present a replication of their main results on samples created through nearest neighbor matching (NNM). We conduct a further replication of this matching analysis using entropy balancing (EB) and coarsend exact matching (CEM). CEM is a restrictive matching method that matches observations from the treatment and the control groups that have identical values on the coarsened variables, as outlined by Iacus et al. (2012). While this method is among the more robust matching algorithms, it does not leave room for much flexibility, as cases that are not matched are dropped. EB is a multivariate reweighting method introduced by Hainmueller (2012). We argue that EB has several advantages over NNM and CEM, including a better covariate balance and the ability to retain most of the original sample size (bar those observations with missing values on the relevant covariates). Instead of choosing the closest units between the treatment and the control group to be matched with one another (as is done in NNM), every unit in the control group is assigned a weight based on entropy balancing, so that perfect covariate balance between the treatment and control group is achieved. The three relevant covariates that we use for balancing are the gender, education, and hometown (whether a respondent lives in a big city or not) of the respondents.

As can be seen in Table 1, the results are robust to EB and CEM and quite similar between the analyses on the unmatched sample, the unmatched sample excluding NAs, the NNM sample, the CEM sample, and the EB sample. In all five cases, the effect of losing the ability to vote on external efficacy and on satisfaction with democracy is statistically significant, while the effect on internal efficacy and political interest is not significant. However, when using CEM, the effect on democratic satistfaction is only significant at the 10% level (p = 0.054). Figure 1 also shows the coefficients with 95% and 90% confidence intervals for the effect of regaining the ability to vote and the net effect of temporary disenfranchisement on the four outcomes of interest when using the EB and CEM. Again, the results lead to the same conclusions as in the original analyses.

Table 1: DiD coefficients (and standard errors) for the effect of **losing eligibility to vote** on the four outcomes, using five different matching schemes: Unmatched, Unmatched excluding NAs, Nearest Neighbor Matching, Coarsened Exact Matching, Entropy Balancing. Note: all coefficients and standard errors are rounded to the second decimal place.

	Unmatched	Unmatched (No NA)	NNM	CEM	EB
Ext.Eff.	-0.19**	-0.20***	-0.23***	-0.19**	-0.19**
	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)
Dem.Sat.	-0.14**	-0.13**	-0.14**	-0.13*	-0.14**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Int.Eff.	0.03	0.04	0.05	0.04	0.03
	(0.09)	(0.09)	(0.11)	(0.09)	(0.09)
Pol.Int.	0.03	0.02	0.03	0.02	0.03
	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)

p-value < 0.01: ***
p-value < 0.05: **
p-value < 0.1: *

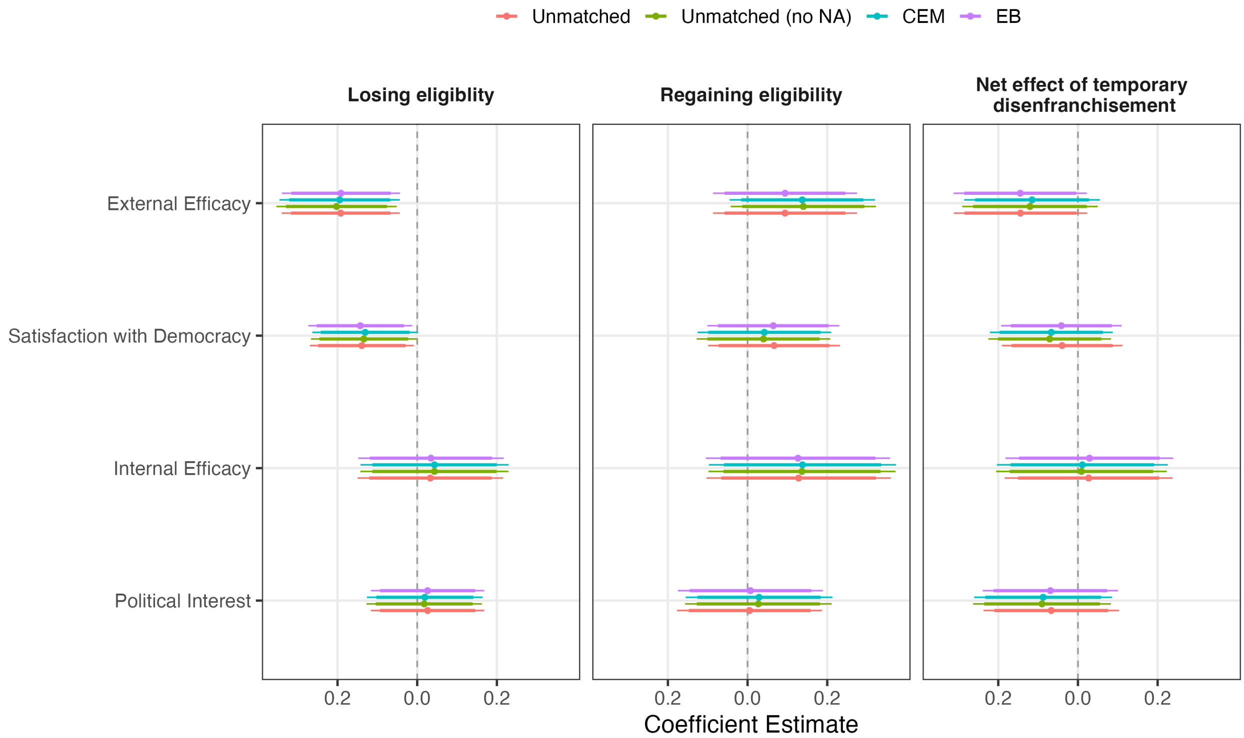

Figure 1: Replication of Figure 2 by Leininger et al. (2023) using Entropy Balancing, Coarsened Exact Matching, and NA exclusion. Note: the error bars are 95% and 90% confidence intervals.

4.2 Panel Attrition

In the discussion of their research design, Leininger et al. (2023) note that panel attrition is ”very similar” between treated and untreated respondents. However, when looking at Table B3 in their Appendix, depicting attrition rates across groups and panel waves, the difference in panel attrition from wave 1 to wave 2 between respondents in group 1 (untreated) and group 2 (treated) seems quite sizeable: while roughly 54% of the untreated respondents responded to wave 2, only roughly 47% of the treated respondents did. Using a t-test on the difference in proportional responses to wave 2 between these two groups, we show that this difference is statistically significant (t = -3.54). This could potentially reduce the size of the treatment effect if those treated respondents who feel more disenfranchised are less likely to respond to wave 2 or inflate it if those who feel more disenfranchised are more likely to respond. Leininger et al. (2023) do not address this, other than reducing the analysis sample to those respondents that participated in both waves. This approach only leads to unbiased treatment effect estimates under the assumption that the ATE would have been the same for those respondents that dropped out of the sample.

Table 2: Results of t-tests to assess differential attrition between treated and untreated respondents and its potential consequences on covariate distributions. Note: all values are rounded to the second decimal place.

T-Test	Difference [95% CI]	T-Value	P-Value
Difference in proportion of wave 2 respondents between treatment and control group	-0.07 [-0.10, -0.03]	-3.54	0.00
Difference in gender between wave 2 respondents and dropouts	0.03 [-0.01, 0.07]	1.68	0.09
Difference in education between wave 2 respondents and dropouts	0.20 [0.15, 0.24]	8.00	0.00
Difference in city size between wave 2 respondents and dropouts	0.01 [-0.03, 0.04]	0.51	0.61

To assess whether those that dropped out of the panel before wave 2 are systematically different from those that remained, we also conduct t-tests of covariate balances between dropouts and remainers. Table 2 reports the results of all t-tests. As can be seen, those study participants that responded to panel wave 2 are significantly more educated than those that dropped out (1% level) and significantly more often female (10% level). There is no difference in the proportion living in big or small cities between wave 2 respondents and dropouts. In sum, these results suggest that there is significant differential attrition between the treatment and the control group, and that those dropping out before panel wave 2 differ significantly from those remaining in the panel on several attributes. The precise degree to which this biases the estimated treatment effects is fundamentally untestable due to unobservables.

Table 3: Coefficients and standard errors for diff-in-diff analyses on highly and less educated subsamples. Note: all coefficients and standard errors are rounded to the second decimal place

	Ext Eff / High Ed	Ext Eff / Low Ed	Dem Sat / High Ed	Dem Sat / Low Ed
Treatment	0.04	0.0	-0.01	0.10
	(0.06)	(0.14)	(0.05)	(0.13)
2nd Wave	0.02	-0.21	-0.03	0.15
	(0.06)	(0.17)	(0.06)	(0.15)
Treat*2nd	-0.24***	0.03	-0.09	-0.36*
	(0.08)	(0.21)	(0.07)	(0.19)
Intercept	3.12***	2.81***	3.81***	3.33***
	(0.04)	(0.11)	(0.04)	(0.10)
N	2311	452	2319	457

p-value < 0.01: ***
p-value < 0.05: **
p-value < 0.1: *

However, when approaching this bias by conducting diff-in-diff analyses on subsamples of highly and less educated respondents (as the two subgroups with the most significant differential attrition), we can see that the effect of losing the eligibility to vote on external efficacy is negative (and larger than in the main analysis) for the highly educated respondents, but positive (and close to zero) for the less educated respondents. The effect of losing the eligibility to vote on democratic satisfaction on the other hand is negative for both subgroups but has a far larger negative coefficient for less educated respondents. This suggests that differential attrition (with less educated respondents dropping out of the panel before wave 2 significantly more frequently than highly educated respondents) leads to an overestimation of the treatment effect on external efficacy, but an underestimation of the treatment effect on democratic satisfaction. These results are reported in Table 3. As a note of caution, the standard errors between the subgroup analyses are hardly comparable, because the subsample of highly educated respondents is much larger than the subsample of less educated respondents.

5 Conclusion

This replication paper attempted to assess the computational reproducibility and robustness replicability of the Leininger et al. (2023) paper. In the original paper, the authors hypothesize that inconsistent voting eligibility laws leading to temporary disenfranchisement across different elections may have negative effects on attitudes related to satisfaction with the political system.

In particular, the authors expect temporary disenfranchisement to depress feelings of external efficacy and satisfaction with democracy but leave internal political efficacy and political interest unaffected. Through a panel study in Germany and a difference-in-differences empirical setup, the authors find support for their theoretical expectations. What's more, regaining eligibility for the next elections does not appear to be associated with statistically significant increases in any of the considered measures, and the net effect of disenfranchisement (and reenfranchisement) remains negative in the case of external efficacy. The authors also conducted several robustness checks, as described in the original paper's appendices.

Beyond the successful computational reproducibility test, we extended those robustness checks trough entropy balancing (EB) and coarsened exact matching (CEM). Our juxtaposition of the new (balanced) coefficients to the corresponding estimates of the original study reveals sign and significance congruence with the slight exception of the satisfaction with democracy outcome under CEM, and leads to virtually the same substantive conclusions.

Next, we took a closer look into the potential biasing effect of differential attrition rates between control and treatment group. We found a statistically significant difference between second-wave response rates across those groups using a t-test, which could potentially impact the effect size of the treatment. We further document patterns of differential attrition through balance tests and show its potential biasing effect on treatment size through a subsample analysis. We highlight this potential issue, while appreciating that it largely lies outside the control of the authors.

Overall, we might conclude this replication report by noting the robustness of this paper to alternative specifications, and the careful attention toward ensuring computational reproducibility by the original authors.

References

Hainmueller, J.: 2012, Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies, *Political Analysis* **20**(1), 25–46.

Iacus, S. M., King, G. and Porro, G.: 2012, Causal Inference without Balance Checking: Coarsened Exact Matching, *Political Analysis* **20**(1), 1–24.

Leininger, A., Sohnius, M.-L., Faas, T., Roßteutscher, S. and Schäfer, A.: 2023, Temporary disenfranchisement: negative side effects of lowering the voting age, *American Political Science Review* **117**(1), 355–361.