INSTITUTE for REPLICATION

No. 99
I4R DISCUSSION PAPER SERIES

Replicating "Run-off elections in the laboratory"

Carina I. Hausladen

Shiang-Hung Hu

Joel M. Levin

January 2024

I4R Discussion Paper Series

I4R DP No. 99

Replicating "Run-off elections in the laboratory"

Carina I. Hausladen[1], Shiang-Hung Hu[2], Joel M. Levin[3]

[1]California Institute of Technology, Pasadena/USA, and ETH Zurich/Switzerland
[2]California Institute of Technology, Pasadena/USA
[3]University of Pittsburgh/USA and University of California, San Diego/USA

January 2024

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

Replicating "Run-off elections in the laboratory"*

Carina I. Hausladen[1], Shiang-Hung Hu[1+], Joel M. Levin[1]

January 13, 2024

Abstract

Bouton et al. (2022) compare the properties of majority run-off and plurality rule elections in a laboratory setting, focusing on Duverger's prediction that plurality rule leads to higher levels of strategic voting. They produce a causal estimate of the difference in incidence of strategic voting across systems, finding more strategic voting under the plurality rule. However, they find that coordination is only higher under the plurality rule when voters are sufficiently divided over which candidate they prefer. They conclude that differences in electoral outcomes and voters' welfare are modest.

We are able to computationally reproduce the original study's main findings using the authors' replication package. The replication package contained both raw data and a cleaned dataset, but did not include a script for cleaning the raw data or a codebook to make sense of it. Therefore, the majority of our work focused on producing code to evaluate and clean the authors' raw data. The authors sent a very helpful response to an earlier draft of this report and their communication improved the quality of our replication effort.

Keywords: Majority run-off system, Plurality rule, Duverger's prediction, Voting behavior, Sincere voting

JEL codes: C92 - Laboratory, Group Behavior D70 - General.

*Equal contribution[1], Corresponding author[+]. Carina I. Hausladen, California Institute of Technology, ETH Zurich. E-mail: carinah@caltech.edu. Shiang-Hung Hu, California Institute of Technology. E-mail: shu3@caltech.edu. Joel Levin, University of Pittsburgh and University of California San Diego. E-mail: joel@ucsd.edu. The authors declare no conflict of interest. The authors have no affiliation with the original authors.

1 Introduction

The main data source for the study is derived from a laboratory experiment that took place over the course of 2017 and 2018. The experiment involved international and US university students participating in the subject pool. The study implemented different treatments, including variations in the distribution of preferences (baseline and low disagreement) and alterations to the electoral system (plurality or run-off voting, detailed in Table 1). These estimates apply specifically to the population of university students involved in the experiment during the specified time period.

The paper examined the impact of different electoral systems on strategic voting and coordination within the experimental population. The authors employed random effect models with clustered standard errors at the group level to analyze the data.

The main results indicate that, for the baseline treatment, there was a significant effect of 16.4% (SE: .0964, $p<.1$), representing the difference in the share of sincere voting under the run-off system compared to the plurality system. In contrast, for the low disagreement treatment, the effect was non-significant at -.597% (SE: .0568, $p>.1$). These percentages illustrate the discrepancy in the proportion of sincere voting between the run-off and plurality systems.

In the following, we replicate the claims mentioned in the abstract and the main tables and figures in the main text.

We accessed the data and codes for the paper from the Economic Journal website. However, one of the authors initially had difficulty finding the data and requested it from the original paper's authors. The authors promptly responded and provided the complete replication package.

To ensure transparency and accessibility, we have stored the data and code used for replication in a public GitHub repository: https://github.com/carinahausladen/runoff-elections.

During the replication process, we successfully replicated the tables and figures by re-executing the Stata code provided by the original authors. We thoroughly examined the code and did not identify any coding errors or discrepancies.

Additionally, we ran the original authors' code on our version of the "cleaned data." We obtained this cleaned dataset by applying our own preprocessing code to the raw data (zTree output) initially provided by the original authors. We compared our cleaned dataset to the dataset used by the authors and found consistency between the two.

Overall, our replication efforts confirm the validity and reliability of the original study's findings.

2 Replication

We tested the study's computational and direct replicability. While the authors provided the raw data, they did not include a script for cleaning the data. Consequently, we developed a data cleaning script to ensure accurate and consistent data processing.

Successfull replication Concerning computational replicability, we reproduced the three main Tables based on the cleaned data and the Stata code provided by Bouton et al. (2022). We did not find any difference between our reproduced tables Table 5, Table 6, Table 7 and the tables printed in Bouton et al. (2022).

Furthermore, we replicate the central figure from the study. This is achieved by generating Figure 1 based on the provided, clean data and juxtaposing it with Figure 2, which is based on our curated data. Our findings show an indistinguishable similarity between both illustrations.

Consequently, it can be concluded that replicating the original figure and tables has been achieved successfully.

Unsuccessful turned into successful replication Initially, we struggled to replicate a study figure due to using our curated dataset rather than the author-provided cleaned one.

The authors identified a coding error in our file and assisted us in refining it. Specifically, the issue was traced to the 'fill' function used for filling missing socio-demographic values. This function filled missing values by using the closest non-missing value in the same group. However, a glitch in the zTree output files assigned a value of 0 to the socio-demographic variables for the first row of each session. This incorrect value was then propagated to other rows due to the 'fill' function. To rectify this, we replaced the values of the socio-demographic characteristics for the first period of each session with missing values before applying the fill function. This correction ensured that the replicated dataset matched the original dataset, allowing for an accurate replication of Figure 6 and its associated regression tables.

Table 1: Original replication: Individual Characteristics and 'Always Duverger' and 'Always Sincere' Behaviour.

	Replication		**Original**	
	Duvergerian	**Sincere**	**Duvergerian**	**Sincere**
gender	-0.0932 (-1.14)	0.171* (2.82)	-0.167 (-1.81)	0.175* (2.91)
age	-0.00277 (-0.23)	0.00803 (0.99)	-0.0272 (-1.73)	0.0154 (1.31)
year	-0.0108 (-0.34)	-0.000207 (-0.01)	-0.0161 (-0.80)	-0.00169 (-0.18)
risk	0.0276 (0.64)	−0.0661* (2.57)	0.000698 (0.02)	−0.0620* (2.30)
trust	0.0670 (1.25)	-0.0308 (-1.01)	0.0538 (0.95)	-0.0209 (-0.78)
experiments high	-0.0116 (-0.13)	0.0383 (0.76)	-0.00408 (-0.04)	0.0397 (0.78)
politics high	0.0577 (0.67)	-0.0637 (-0.99)	0.0487 (0.57)	-0.0576 (-0.96)
cons	0.253 (2.11)	-0.0212 (-0.91)	1.030* (2.29)	-0.232 (-0.72)
N	132		132	

t-statistics are denoted in parentheses. $*$ signifies $p < 0.05$, $**$ denotes $p < 0.01$, $***$ represents $p < 0.001$.

Additionally, we identified some apparent errors in the structure of Table G3, which describes theoretically derived thresholds across treatments. The treatment indicators (on the left side of the table) are: "P_B, P_LD, P_SM, R_NU." Elsewhere in the text, the first letter in these treatment indicators denotes election rule (P denoting plurality; R denoting runoff), and the second letter(s) denote combinations of parameters (e.g., SM denoting small minority.) In this table, it appears that the authors intended the indicators on the left side of the table to denote only parameters but not election rules, given that there are two columns for election rules. Additionally, this table suggests that there is a threshold for a no-upset

Table 2: Revised replication: Individual Characteristics and 'Always Duverger' and 'Always Sincere' Behaviour.

	Replication revised		Original	
	Duvergerian	**Sincere**	**Duvergerian**	**Sincere**
gender	-0.167 (-1.81)	0.175* (2.91)	-0.167 (-1.81)	0.175* (2.91)
age	-0.0272 (-1.73)	0.0154 (1.31)	-0.0272 (-1.73)	0.0154 (1.31)
year	-0.0161 (-0.80)	-0.00169 (-0.18)	-0.0161 (-0.80)	-0.00169 (-0.18)
risk	0.000698 (0.02)	−0.0620* (2.30)	0.000698 (0.02)	−0.0620* (2.30)
trust	0.0538 (0.95)	-0.0209 (-0.78)	0.0538 (0.95)	-0.0209 (-0.78)
experiments high	-0.00408 (-0.04)	0.0397 (0.78)	-0.00408 (-0.04)	0.0397 (0.78)
politics high	0.0487 (0.57)	-0.0576 (-0.96)	0.0487 (0.57)	-0.0576 (-0.96)
cons	1.030* (2.29)	-0.232 (-0.72)	1.030* (2.29)	-0.232 (-0.72)
N	132		132	

t-statistics are denoted in parentheses. $*$ signifies $p < 0.05$, $**$ denotes $p < 0.01$, $***$ represents $p < 0.001$.

Table 3: Original replication: P-values for Individual Characteristics and 'Always Duverger' and 'Always Sincere' Behaviour.

	Replication		Original	
	Duvergerian	**Sincere**	**Duvergerian**	**Sincere**
gender	0.281	0.018	0.100	0.015
age	0.822	0.344	0.114	0.221
year	0.739	0.992	0.441	0.864
risk	0.540	0.028	0.987	0.044
trust	0.238	0.335	0.365	0.453
experiments high	0.896	0.467	0.966	0.455
politics high	0.517	0.346	0.580	0.359
cons	0.061	0.382	0.045	0.489
N	132		132	

plurality rule condition, which does not exist in their study, and that there is no threshold for the no-upset runoff rule condition, which does exist. Ideally, the manuscript should be corrected to clarify these issues.

3 Conclusion

The replication package provided by Bouton et al. (2022) was well-organized, enabling us to effectively reproduce the study's primary findings. The authors constructed a refined dataset, yet they did not include a script that could transform the raw data—namely the zTree output—into a processed format. Consequently,

Table 4: Revised replication: P-values for Individual Characteristics and 'Always Duverger' and 'Always Sincere' Behaviour.

	Replication revised		Original	
	Duvergerian	**Sincere**	**Duvergerian**	**Sincere**
gender	0.100	0.015	0.100	0.015
age	0.114	0.221	0.114	0.221
year	0.441	0.864	0.441	0.864
risk	0.987	0.044	0.987	0.044
trust	0.365	0.453	0.365	0.453
experiments high	0.966	0.455	0.966	0.455
politics high	0.580	0.359	0.580	0.359
cons	0.045	0.489	0.045	0.489
N	132		132	

we developed our own script for this purpose. This task presented challenges, as a few columns in the cleaned dataset were not explained in the codebook. Thus, we had to infer their meanings based on the data provided and the context given in the text. Despite these obstacles, we successfully interpreted these columns. We encountered some discrepancies when replicating the numerical values of a secondary finding from the study, upon which the authors helped us finding a coding mistake we made. Subsequently, this finding could be replicated successfully.

References

Bouton, L., Gallego, J., Llorente-Saguer, A. and Morton, R.: 2022, Run-off elections in the laboratory, *The Economic Journal* **132**(641), 106–146.

4 Figures

We replicate the central figure from the study. This is achieved by utilizing the provided, clean data, as referenced in Figure 1, and juxtaposing it with the restructured data that we have generated from the primary datasets, as seen in Figure 2. Our findings show an indistinguishable similarity between both illustrations. Consequently, it can be concluded that the replication of the original figure has been achieved successfully.

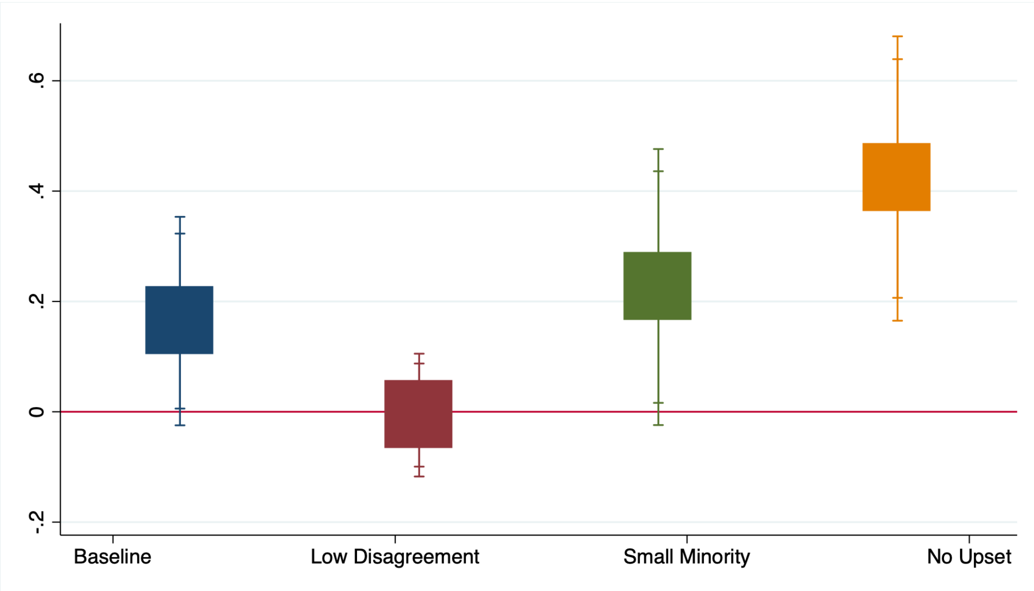

Figure 1: Figure based on provided, cleaned data: Treatment Effect of the Voting Rule on the Probability of Strategic Voting by t_B-Voters: 90% and 95% Confidence Interval (Original Figure 3)

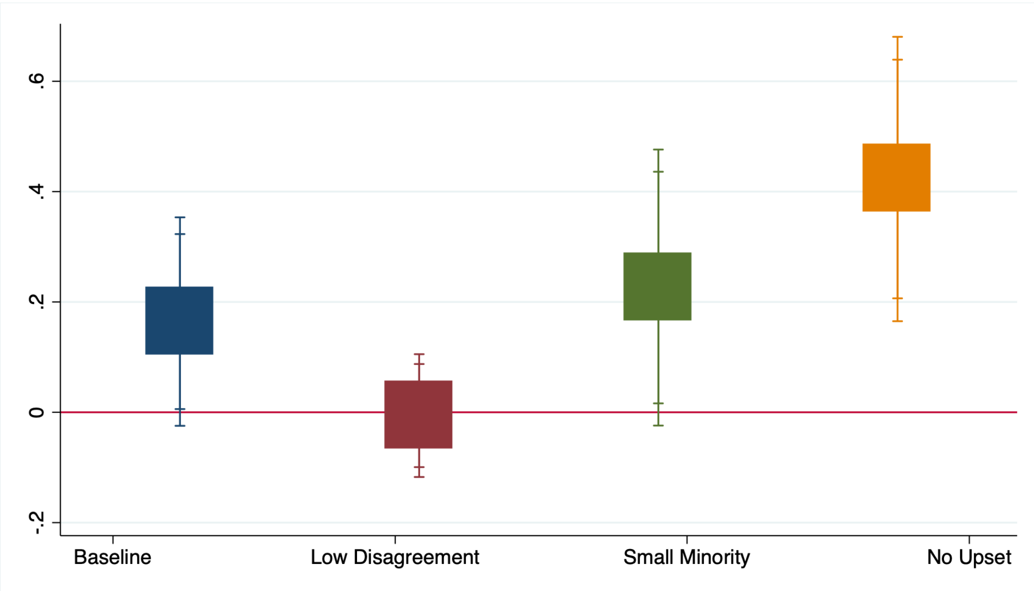

Figure 2: Figure based on data constructed from raw datasets: Treatment Effect of the Voting Rule on the Probability of Strategic Voting by t_B-Voters: 90% and 95% Confidence Interval (Original Figure 3)

5 Tables

The tables presented below represent key data from the original study. They were reconstructed utilizing the supplied cleaned dataset and Stata code. Upon examination, we discovered a perfect alignment in the numerical data between our recreated tables and those published in the original paper. Furthermore, we replicated Table 5 with our constructed dataset and found no difference. We did not replicate Table 6 and Table 7 with our dataset because the tables require simulation.

Table 5: Aggregate Behavior in Main Treatment (Original Table 2)

		baseline			Low disagreement		
		% A	% B	% C	% A	% B	% C
Plurality	t_A	99.1	.41	.49	98.17	1.57	.26
	t_B	75.25	24.14	.62	71.49	27.92	.59
	t_C	1.47	.89	97.64	2.82	1.22	95.97
Run-off	t_A	97.21	2.38	.41	97.34	2.14	.52
	t_B	58.72	40.26	1.01	70.55	28.6	.85
	t_C	1.34	.97	97.69	1.58	.69	97.73

Table 6: Simulated Outcomes in Main Treatments (Original Table 3)

			baseline			Low disagreement		
			% A	% B	% C	% A	% B	% C
Plurality	Realised	All periods	67.69	.09	32.22	70.74	0.0	29.26
		Second half	70.4	.03	29.57	70.67	0.0	29.32
	Theory	Duvergerian eq.	76.3	0.0	23.7	76.27	0.0	23.73
Run-off	Realised	All periods	65.31	1.81	32.88	69.82	.06	0.11
		Second half	66.23	1.23	32.53	69.69	.02	30.29
	Theory	Duvergerian eq.	71.24	0.0	28.76	71.51	0	28.49
		Sincere	50.12	13.13	36.75	-	-	-

Table 7: Simulated Outcomes in Main Treatments (Original Table C7)

	(1) A Victory	(2) B Victory	(3) C Victory	(4) A Victory	(5) B Victory	(6) C Victory
Baseline	0.0433	0.0111**	-0.0544*			
	(0.0334)	(0.00532)	(0.0328)			
Low Disagreement				-0.0267	-	0.0267
				(0.0306)	-	(0.0306)
Control group	P_B	P_B	P_B	P_{LD}	P_{LD}	P_{LD}
Observations	660	660	660	600	600	600

Standard errors are denoted in parentheses. $*$ signifies $p < 0.05$, $**$ denotes $p < 0.01$, $***$ represents $p < 0.001$. The exact p-values from the most left to the most right statistic in order are: 0.195, 0.037, 0.097, 0.383, NA, 0.383.