Salomo Hirvonen, Maarit Lassander, Lauri Sääksvuori, and Janne Tukiainen

Who is mobilized to vote by short text messages? Evidence from a nationwide field experiment with young voters

Aboa Centre for Economics
Discussion paper No. 157
Turku
April 2024 (previous versions: June 2023, January 2023)

The Aboa Centre for Economics is a joint initiative of the economics departments of the University of Turku and Åbo Akademi University.

Copyright © Author(s)

ISSN 1796-3133

Printed in Uniprint
Turku
April 2024 (previous versions: June 2023, January 2023)

Salomo Hirvonen, Maarit Lassander, Lauri Sääksvuori, and Janne Tukiainen

Who is mobilized to vote by short text messages? Evidence from a nationwide field experiment with young voters

Aboa Centre for Economics
Discussion paper No. 157
April 2024 (previous versions: June 2023, January 2023)

ABSTRACT

Using a large randomized controlled trial and rich individual-level data on eligible voters and their household members, we evaluate how get-out-the-vote (GOTV) appeals affect inequalities in voting, transmit from treated to untreated individuals within households, and how the transmission of voting decisions through family networks influences inequalities in voting. We find that receiving a text message reminder before the Finnish county elections in 2022 mobilized mainly low-propensity voters, and thereby reduced existing social inequalities in voting within the target group of young voters. We remarkably find that over 100 percent of the direct treatment effect spilled over to untreated household members. These spillovers reduced inequality in political participation among the older voters that were not part of the target group. Overall, our results exemplify how randomized controlled trials with a limited focus on the analysis of individuals in the treatment and control groups alone may lead to misestimating the compositional effects of get-out-the-vote interventions.

JEL Classification: C93, D72

Keywords: Get-out-the-vote, Field experiments, Spillover effects, Voter turnout

Contact information

Salomo Hirvonen
Department of Economics, University of Turku.

Maarit Lassander
Prime Minister's Office, Finland.

Lauri Sääksvuori
Finnish Institute for Health and Welfare, Finland.

Janne Tukiainen (corresponding author)
Department of Economics, University of Turku.
Email: janne.tukiainen (at) utu.fi

Acknowledgements

We thank the seminar participants at FEAAM, the University of Helsinki, Laurea, and SABE/IAREP for helpful comments. We are grateful to the Ministry of Justice (Finland) and the Prime Minister's Office (Finland) for funding and help. The authors have no relevant financial or non-financial interests to disclose. This RCT was registered in the American Economic Association Registry for randomized control trials (AEARCTR-0008790). The Ethics Committee for Human Sciences at the University of Turku approved this study (48/2021). The data used are proprietary and accessible only through Statistics Finland's remote access system. Thus we are not able to share the data, but researchers can replicate our results by the code provided by us and purchasing the same data sets from Statistics Finland together with acquiring access to Statistics Finland's remote access system.

Who is mobilized to vote by short text messages?

Evidence from a nationwide field experiment with young voters

1 Introduction

Voter mobilization through get-out-the-vote (GOTV) interventions is often successful in increasing voter turnout. However, the effects of mobilization campaigns may not be evenly distributed across the electorate and can potentially increase inequalities in political participation (Enos et al., 2014). As a result, GOTV interventions, even with the best of intentions, may aggravate uneven participation in voting and exacerbate disparities in descriptive and substantive representation that are biased against underprivileged citizens (Fowler, 2013; Harjunen et al., 2023).

The effects of GOTV interventions on voter turnout have been examined in numerous experimental studies that have covered various mobilization tactics, explored different means of communication (e.g., door-to-door canvassing, postcard mailings, phone calls and text message reminders) and investigated the potential behavioral mechanisms driving the impact of mobilization campaigns on voter turnout. In a standard randomized controlled trial (RCT), a researcher assigns individuals or households to receive a treatment before the elections and estimates the average treatment effect on voter turnout by comparing the differences in turnout between the treatment and control groups after the elections. Since the early GOTV experiments, randomized controlled trials have proven to be a feasible method to evaluate the effectiveness of GOTV appeals across different electoral contexts, enabling researchers to accumulate a large body of evidence about the impact of various mobilization interventions on voter turnout. However, there still exists limited knowledge how the effects of GOTV interventions may spill over from treated individuals to untreated individuals and how this transmission of voting decisions in social networks may affect the composition of the electorate.

This paper evaluates how GOTV appeals affect inequalities in voting, transmit from treated to untreated individuals within households, and how this transmission of voting decisions through family networks affects inequalities in voting. For this purpose, we conduct a large RCT and evaluate the effectiveness of non-partisan text message reminders on voter turnout in nationwide county elections in Finland. The target population of our experiment are young adults aged between 18 and 29 years, a population group with high levels of education and health in a global comparison, but persistently low turnout rates.[1]

The main contribution of this article is the measurement of spillover effects and, more importantly, the examination of heterogeneous spillover effects by voting propensities. First, using electronic voter turnout records and rich individual-level administrative data on eligible voters, we estimate the effect of a large non-partisan text message-based GOTV campaign on social inequalities in voting. Second, using unique household IDs, we investigate how turnout decisions transmit between household members. Finally, we examine how large spillovers from treated individuals to untreated individuals may influence the effects of GOTV mobilization campaigns on the composition of the electorate.

Our estimation of the electoral effects builds upon access to administrative records that contain comprehensive information about individuals' own and their household members' demographic background and labor market outcomes. Thus, the linked dataset contains a barrage of potential individual-level characteristics that could be used to explore the heterogeneity of direct and spillover effects among the eligible voters. To keep our statistical analyses conceptually tractable and facilitate the interpretation of results, we organize the estimation of heterogeneous treatment effects into two categories. First, following the paper by Hirvonen et al. (2023), we examine the potential heterogeneity of treatment and spillover effects using data-driven estimation methods. This exploration of heterogeneous effects contributes to the literature on voter mobilization and inequalities in political participation. These results, however, may hold limited practical implications for the development of concrete targeting strategies to ameliorate the stubborn demographic gaps in political participation. Thus, we complement the analysis by estimating heterogeneous treatment effects in various (pre-registered) sub-samples that are created using a single observed characteristic at a time. Through these complementary analyses, we aim to advance the applied research literature that assesses the potential of enhancing the effectiveness GOTV interventions through selective targeting of interventions.

Our paper relates to several strands of literature. First, our study relates to the literature on social inequalities in political participation. To date, there is relatively little evidence on the compositional

[1] A systematic assessment of expected human capital formation for children born in 195 different countries ranks Finland as the country with the highest level of expected human capital in the world (Lim et al., 2018). Despite the high levels of human capital among the young adults, Finland has one of the largest age gaps in turnout between older (aged 60 and above) and younger (aged from 18 to 29 years) voters (Mo et al., 2022). Human capital is characterized in this context as the aggregate level of education, training, skills, and health in a population.

effects of text message-based mobilization strategies. A recent literature on the compositional effects of get-out-the-vote mobilization strategies suggests that many current mobilization strategies may widen existing social disparities in voting by predominantly mobilizing high-propensity voters instead of under-represented low-propensity voters (Enos et al., 2014). Simultaneously, there is evidence that the success and compositional effects of GOTV mobilization strategies may vary by the salience of the elections (Arceneaux and Nickerson, 2009). Our paper complements the existing literature on the compositional effects of GOTV mobilization strategies in several ways. First, and most importantly, we assess how accounting for spillover effects from treated individuals to untreated individuals modifies the effect of GOTV mobilization on the composition of the electorate. Second, we assess the compositional effects of GOTV mobilization in an electoral context where all eligible citizens are automatically registered to vote. Third, we assess the robustness of the prevailing empirical strategy in the relevant literature that estimates baseline voting propensities using within-sample covariates and interacts the predicted propensities to vote with the GOTV treatment indicator. To address the concern that the within-sample estimates of voting propensities may not predict turnout out-of-sample, we estimate citizens' propensities to vote using machine learning techniques. Consistent with the existing Nordic GOTV interventions (Bhatti et al., 2018; Bergh et al., 2020; Bergh and Christensen, 2024), we find that in particular low-propensity voters are mobilized. More uniquely, we find that the effects are the largest for the citizens belonging to low vote propensity groups, but who have high personal interest in voting.

Second, our paper relates to the very few experimental studies on voter mobilization with an explicit objective to measure spillover effects. Prior to our work, Nickerson (2008), Sinclair et al. (2012) and Bhatti et al. (2017a) have investigated how the effects of different get-out-the-vote appeals may transmit from treated to untreated individuals and reported within household spillover effects varying from 30% to 60% of the direct effect. We find that over 100% of the direct treatment effect spilled over to the untreated household members. Moreover, our findings complement the existing literature on the measurement of spillover effects by estimating how spillovers interact with the eligible voters' predicted propensity to vote and affect the composition of the electorate. We find that also the spillover effects reduce the existing inequalities in voting. These results stress that an inadequate analysis of the spillover effects may not only lead to severely underestimate the effectiveness of GOTV interventions in general, but also to overlook the effects of GOTV mobilizations on the composition of the electorate.

Our results related to the direct effect of the intervention on voter turnout are largely consistent with the existing Nordic literature that has systemically, although with varying magnitudes, documented the effectiveness of SMS reminders on voter mobilization (Bhatti et al., 2017a,b; Bergh et al., 2021; Naess, 2022; Bergh and Christensen, 2024). We find a statistically significant, about 0.9 percentage point, direct average treatment effect in the probability of voting. The effect size of 0.9 percentage points equals 3% increase compared to the control group average turnout of 30.9%. Our evidence stems from a relatively low-salience elections and may not readily generalize to high-salience elections (Malhotra et al., 2011; Green and Gerber, 2019; Bergh and Christensen, 2024; Mann and Haenschen, 2024).

Finally, our study builds on and contributes to the literature that has investigated the effectiveness of numerous voter mobilization strategies and different contents of campaign messages on voter turnout and choice (Green et al., 2013; Green and Gerber, 2019). We find the largest effects for neutral message types.

The paper proceeds as follows. In Section 2, we discuss theoretical arguments regarding the potential effectiveness of the intervention. Section 3 describes the relevant electoral context. In section 4, we describe our data, experimental design and the sample. Section 5 presents our empirical methods. Section 6 presents the results and Section 7 discusses them before Section 8 concludes.

2 Theoretical background

The most prominent theory regarding the average treatment effects of text message reminders on turnout is the Noticeable Reminder Theory (Dale and Strauss, 2009; Malhotra et al., 2011). It's premise is that registered voters tend to have an intention to vote, but may fail to do so, because of time constraints and lack of planning and attention, in which case only a simple nudge (Thaler and Sunstein, 2009) is needed to remind them on their intention to vote, and thus, mobilize them. According to Dale and Strauss (2009), text messages are likely to surpass the threshold of attention, because voters pay attention to their phones and text messages are difficult to ignore. However, in our context, all eligible voters are automatically registered, and thus, this theory does not perfectly apply. Nonetheless, a substantial share of eligible voters are likely to have the intention to vote and yet fail to do so.

The Receive-Accept-Sample theory (Zaller, 1992; Matland and Murray, 2012) can explain why different types of potential voters may be differently affected by the experiment. The theory implies that politically more aware citizens who are more likely to vote are receiving more political messages in general, and therefore, less likely to accept any individual message due to the flood of them. Less aware individuals receive fewer messages in general, and thus, are more likely to accept them. This would imply that the effect of an intervention is larger for the inattentive citizens who are less likely to vote. However, the standard assumption in GOTV experiment literature is that inattentive citizens with little interest in the election are unlikely to take note of such a small nudge as a text message. Thus, the effects are expected to be the strongest among the voter group in the middle of the distribution who pay some but not too much attention to politics. This argument is similar to the proposal by Arceneaux and Nickerson (2009) which suggests that citizens who are nearly indifferent between voting or not voting are the easiest to mobilize.

We can hypothesize about the spillover effects using four related theoretical arguments. First, according to the Social Occasion Theory (Dale and Strauss, 2009), if the text message creates communication within the household it creates a social occasion that may establish feelings of connection to voting. Second, social norms and conformity pressure may lead to similar behaviors within households (Zuckerman, 2005; Foos and de Rooij, 2017). Thus, if a treated family member is affected by the experiment, then other members may conform. Third, information transmitted within households is often of low-costs and citizens may pay more attention to information coming from close social network (Klofstad, 2007). The fourth rationale for spillover effects is that many voters choose to vote at a polling station on the Election Day instead of voting in advance and households often go to vote together. The theories two to four imply that the spillover effect can arise only if there is a direct effect, and therefore, the heterogeneity of the spillover effects should be contingent on the heterogeneity of the main effect. However, the Social Occasion Theory allows for the possibility that there is a spillover effect even without the main effect as long as the treatment is discussed among the household members.

3 Background and Context

We conducted our RCT in Finnish nationwide county elections held on January 23, 2022. Counties are the mid-tier level of decision-making in Finland between the municipalities and the central government. They resulted from a recent large social and healthcare reform. Thus, the elections were the first of their kind in Finland. The elections were expected to be of low salience and interest. This expectation also turned out to be true as turnout in the elections was 47.5%, which is lower than in any parliamentary elections in the Finnish history.

The allocation of seats in the county elections is proportional to the votes following d'Hondt system of open party list proportional representation (PR) and identical to the Finnish parliamentary and municipal elections. Finland uses a very pure form of open-lists in the sense that personal vote is obligatory: each voter gives exactly one vote to one candidate. Parties are assigned seats based on the sum of its candidates' personal votes and the seats within the party are assigned purely based on the personal votes. Moreover, candidates are almost always presented in alphabetical order in the ballot lists limiting parties ability to signal their preferences over the candidates. Overall, the open list electoral system in Finland is highly personalised, which may increase incentives for individual campaigning compared to several democracies with closed list PR or mixed electoral systems (von Schoultz and Strandberg, 2024).[2]

Voters are automatically registered in all elections in Finland. An electronic register of all eligible voters (voting register) is established based on the Population Information System on the 46th day before the election day (Jääskeläinen, 2020). All voters listed in the voting register receive a notice of their right to vote (polling card) no later than 24 days before the election day. The polling card indicates the date of the election, the period for advance voting, the locations of advance polling stations within the voter's electoral district, the address of the voter's election day polling station, and contact information of the electoral authorities. The polling stations have only an administrative role as the elections are held at-large in the whole county. A typical characteristic of the Finnish elections is that a relatively large share of voters cast their ballots at polling stations during the period for advance voting that begins 11 days before the election day and ends five days before the actual election day. In the 2022 county elections, 57% of individuals who voted used the advance period to cast their vote.

Prior to our study, text message-based mobilization experiments have been conducted in the US, Denmark and Norway. The Finnish electoral system and voter mobilization environment closely resembles the other Nordic countries. Turnout in Finnish local and regional elections is typically markedly higher than in the local US elections, but has been in many recent elections noticeably lower than in comparable Danish and Norwegian elections (Bhatti et al., 2017b; Bergh et al., 2021). In the 2021 municipal elections, the turnout of eligible voters was 55.1%. There are notable demographic inequalities in voting. Young adults aged from 18 to 29 years are markedly less likely to vote than the older age cohorts. Their turnout in the 2021 municipal elections was 36.6%. The gender gap among young voters in the 2021 municipal elections was 8 percentage points as women had a turnout of 40.7% and men had a turnout of 32.7%.[3]

Voters' access to information on party platforms and individual candidates is supported through wide-ranging public information campaigns and strong public media presence. Political campaigning and advertising is regulated by the Election and Data Protection Acts that restrict the use of personalized advertising using direct mailings, phone calls and text messages. To our knowledge, prior to this study,

[2]In contrast, in the other Nordic countries, parties have a larger role in the electoral system. Sweden nominally uses a flexible list where it is possible to give personal votes. However, a large number of those are needed to change the otherwise closed list. In Norway, municipal elections use open list, but parties can give large personal vote bonuses to their preferred candidates.

[3]The most recent parliamentary elections before the county elections were held in year 2019. The age gap between the young voters (aged from 18 to 29) and older voters (over 29 years of age) in these these elections was around 11 percentage points, and the gender gap within the group of young voters was around 10 percentage points. The gender gap persists almost as large even after controlling for education.

there has not been politically motivated or government sponsored non-partisan text-message campaigns to mobilize voters in Finland.

4 Experimental Design and Data

4.1 Sample

To conduct the experiment, we accessed the electronic register of eligible voters maintained by the Finnish Digital and Population Data Services Agency. This electronic register contains information on voters (e.g., name, personal identity code, electoral district, and the municipality of residence) as recorded in the Population Information System. Importantly, the electronic voting register enables us to link different treatment arms to individual-level electronic records on turnout. Our sample includes municipalities where voting districts having an electronic voting register cover at least 80% of the eligible voters in the municipality. This leads to a sample of 99 municipalities with full electronic voting registry coverage and 19 municipalities with more than 80% coverage out of 309 municipalities. Table 1 shows that 56% of all eligible voters aged 29 years and under live in these municipalities.

After extracting relevant personal information of all eligible voters aged from 18 to 29 years and residing in the voting districts covered by the electronic voting register, we contracted with an IT-company that conducted a search to provide the cell phone numbers of individuals included in the electronic voting register. The matching of eligible voters' personal information to valid cell phone numbers led to an analysis sample of 51101 individuals aged from 18 to 29 years of age.[4]

Table 1 shows descriptive statistics for various samples. Column (1) shows the analysis sample that was used to randomize individuals into treatment and control groups. Column (2) drops from this analysis sample the 18 year old eligible voters to facilitate more accurate comparison between Columns (2) - (4). Column (3) describes all 19-29-year-old individuals contained in the electronic voting register. Column (4) contains the full population belonging to same age cohorts. As we have information only on the year of birth, and not on the exact date of birth that would be necessary for identifying 18-year-old eligible voters from the full population, Columns from (2) to (4) do not include any 18-year-old individuals. By comparing Columns (3) and (4), we find that the demographics in municipalities used to draw our sample due to the availability of the electronic voting register closely resemble the demographics of full equally aged population in Finland. By comparing Columns (2) and (3), we find that the final analysis sample

[4]The company that was used to provide the phone numbers was able to find cell phone numbers for 18.2 percent of individuals included in the electronic voting register.

closely reminds the population of the same age living in the same municipalities with a somewhat lower share of females and immigrants. By contrast, the taxable income is somewhat higher in our analysis sample than in the population of the same age at large. Overall, the comparison of our analysis sample to the full population sample suggests that the restriction to municipalities with an electronic voting register and the loss of individuals because of not observing their phone numbers does not substantially affect the representativeness of our results.

Table 1: Summary statistics: Sample compared to population

	Analysis sample Full Sample (1)	Analysis sample Aged 19 to 29 (2)	Analysis Municipalities Aged 19 to 29 (3)	Full population Aged 19 to 29 (4)
Female	0.40	0.41	0.48	0.49
	(0.49)	(0.49)	(0.50)	(0.50)
Age	24.62	24.65	24.19	24.28
	(3.15)	(3.13)	(3.15)	(3.12)
High School Degree	0.44	0.44	0.44	0.45
	(0.5)	(0.5)	(0.5)	(0.5)
Taxable Income	158781	15808	13539	13972
	(13163)	(13160)	(12399)	(12553)
Immigrant	0.04	0.04	0.07	0.07
	(0.20)	(0.20)	(0.26)	(0.25)
Observations	51,101	50,899	280,925	496,042

Notes: Standard deviations in parentheses. Covariates are measured in year 2019 with the exception of age which is for year 2022. Number of observations for taxable income are 47,503 (Column 1), 47,416 (Column 2), 258,065 (Column 3) and 458,604 (Column 4).

4.2 Experimental design

The experiment was conducted in collaboration with the Ministry of Justice (Finland) and the Prime Minister's Office (Finland). They also funded the experiment. The objectives of our RCT and a study protocol was pre-registered in the American Economic Association Registry for randomized controlled trials as AEARCTR-0008790. The Ethics Committee for Human Sciences at the University of Turku, Finland, approved this study (decision number: 48/2021).

To estimate the direct causal effect and potential spillover effects of alternative text message reminders on voter mobilization, we randomized all individuals in our analysis sample into control and treatment groups. There were three different treatment groups that varied the wording of text messages. We used an allocation ratio that assigned 40 percent of individuals into a control group and 60 percent of individuals into three equally sized treatment groups (Figure 1). We stratified the randomization by municipality to guarantee that 60% of all eligible 18 to 29-year-old voters received a reminder in each municipality. The stratification by municipality is expected to increase the precision of estimated treatment effects (Duflo et al., 2007) and enables us to provide more reliable estimates for local level analyses. At the time of randomization, we did not possess data on other covariates suitable for stratified randomization.

Following the timing of polling opportunities in the Finnish elections, we sent two text messages for all individuals in treatment groups. The first message was sent a day before the beginning of the advance voting period. The second message was sent a day before the election day. There was no variation in the intraday timing of text messages. All messages were simultaneously sent at 4 pm using a mass text messaging service.

We measure the effect of SMS reminders on voter turnout using individual-level data recorded in an electronic register of turnout. The electronic voting record contains a unique identifier for each citizen and a variable indicating whether the person voted in the election. Using unique personal identifiers and household IDs, we merge the voting records with the treatment assignment, comprehensive socio-economic data and pre-existing turnout data that covers citizen's participation in all nationwide elections since 2015. Crucial to the treatment heterogeneity analyses, we are able to merge the voting records with individual-level data on prior voting histories and rich personal information including, among other information, data on voter's labor income, capital income, social transfers, education, ethnicity and employment records. The resulting data are protected from improper disclosure and accessible only through Statistics Finland's remote access system. Thus, we are not able to share the data, but all results can be replicated using the code provided by us and purchasing the mentioned data sets together with acquiring access to Statistics Finland's remote access system.

4.3 Message contents

Since the popularization of the nudge theory (Thaler and Sunstein, 2009), there has been a large influx of studies testing the effectiveness of varying message contents for multiple purposes in numerous different contexts. While there are some broadly heralded examples of cases in which small variations in message contents have led to meaningful differences in behavioral outcomes, in the context of voter mobilization the Noticeable Reminder Theory (Dale and Strauss, 2009) implies that the content of text messages should not affect turnout. However, there is still not much empirical research testing how text message reminders with different types of appeals impact the likelihood of being mobilized to vote.

In addition to examining the overall causal effect of text message reminders on voting, we test the effectiveness of different message contents. For this purpose, we developed three different types of messages that appeal to different motivations to vote. The first type of message was a neutral message that just briefly informed recipients about the forthcoming elections and abstained from expressive and

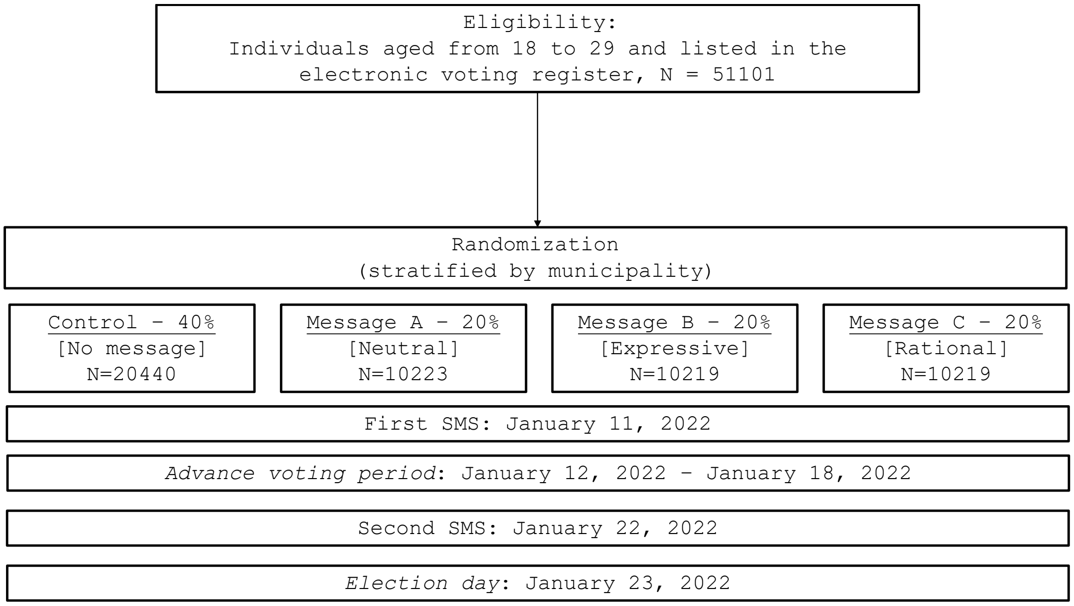


Figure 1: Eligibility, Randomization and Treatment.

instrumental motivations to vote. The second type of messages was developed to appeal to the expressive motive of voting (Brennan and Brooks, 2013) and highlighted voters' right to express their voice by voting. The third type of messages was developed to appeal to a more instrumental or rational motive of voting (Downs, 1957; Lyytikäinen and Tukiainen, 2019) and emphasized recipients' chance to influence the direction of policies and provision of public services through voting. Table 2 shows wording of the messages.

There are conflicting theoretical arguments for whether message type would matter. Many scholars argue that the civic duty component is more important than the instrumental component in the calculus of voting, and therefore, content that refers to civic duty would have larger effect than the instrumental component (Blais et al., 2000; Gerber et al., 2008). On the other hand, many citizens are averse to paternalism in other contexts Lassen and Mahler (2023), and thus, we propose that non-neutral message content may be considered paternalistic, which implies the possibility that neutral message has the largest effect.

All message contents were developed by the authors in collaboration with the electoral authority (Ministry of Justice, Finland) to ensure that the contents conformed with the existing electoral code of conduct. All messages included a hyperlink to a homepage www.vaalit.fi [www.elections.fi] maintained by the electoral authority to provide reliable and unbiased information about the organization of elections in Finland. The electoral authority served as the sender of the messages which is likely to have increased the credibility of messages and set the notifications apart from standard promotional messages that individuals may receive to their phones.

Table 2: Overview of message contents by treatment

Treatment	Message #	Message text
Neutral	#1	”Hi, please remember that county elections will be held on January 23. Domestic advance voting period is from January 12 until January 18. More information at vaalit.fi. Regards, Ministry of Justice.”
Neutral	#2	”Hi, please remember that county elections will be held on January 23. More information at vaalit.fi. Regards, Ministry of Justice.”
Expressive	#1	”Hi, please remember to use your right to vote in country elections on January 23. Domestic advance voting period is from January 12 until January 18. More information at vaalit.fi. Regards, Ministry of Justice.”
Expressive	#2	”Hi, please remember that county elections will be held on January 23. Democracy needs your voice, please use your right to vote. More information at vaalit.fi. Regards, Ministry of Justice.”
Instrumental	#1	”Hi, have your say on community services in county elections on January 23. Domestic advance voting period is from January 12 until January 18. More information at vaalit.fi. Regards, Ministry of Justice.”
Instrumental	#2	”Hi, please remember county elections on January 23. By voting, you can have a say on the organization of health and social care services, and fire and rescue care. More information at vaalit.fi. Regards, Ministry of Justice.”
Control	-	[None]

5 Estimation methods

Given the randomization procedure and access to administrative data containing unique personal and household IDs, we estimate direct effects, spillover effects and treatment effect heterogeneity.

5.1 Direct effects

To assess the overall direct effect of SMS reminders on turnout, we estimate the pooled average treatment effect of receiving any type of reminder in contrast to the counterfactual of receiving no reminder. Moreover, to investigate the direct effect of different contents of reminders on turnout, we estimate the average treatment effects by treatment. As pre-registered, we estimate the direct treatment effects using a linear probability model and progressively add control variables to the model:

$$Y_i = \beta_0 + \beta_1 Treatment_i + \boldsymbol{X}'_i \boldsymbol{\beta} + \epsilon_i,$$

where $Treatment_i$ indicates treatment assignment and $X_i'\boldsymbol{\beta}$ individual level demographic controls.[5] Our demographic controls include educational background, which is defined as the mother of the individual having a high school degree or using individual's own high school degree status if our data does not allow us to identify the mother of the individual (29% of our sample) based on the household data going back to year 2011.[6] In addition to the educational background, we use logarithm of individual's mother's taxable income and occupation as controls for the socio-economic background. As our sample consists of young voters, we believe that their mothers' characteristics are more accurate in describing individuals circumstances and predicting voting than their own characteristics. In addition to educational and socio-economic background, we include individuals' ethnicity, which takes value 1 if person's both parents are born outside of Finland. We also include age, gender and an indicator variable documenting if the individual was eligible to vote in the 2022 elections for the first time. Adding control variables to the estimations of average treatment effects in a randomized experiment is not expected to affect the point estimates, but can reduce residual variance and increase the precision of the estimates. We cluster standard errors at the municipal level.[7]

5.2 Spillover effects

Unique household IDs included in our data enable us to investigate spillover effects of our get-out-the-vote intervention within the households.[8] To study the intra-household transmission of treatment effects after receiving an SMS reminder, we restrict our sample to households where there was either exactly one young voter who was part of the treatment group or there was exactly one young voter who was part of the control group, leading to a sample size of 51.4% of the total sample as a high proportion of individuals in our sample are living alone. Thus, we drop from the spillover estimation sample households where there are more than one potentially treated young voter.[9] Therefore, the treatment group for spillover

[5]Following our pre-analysis plan, we conduct supplementary analyses using Logit models to study the robustness of our linear probability model estimates. Results from these estimations are reported in the Online Appendix (Tables A1 and A2) and show that our results are robust to the choice of the estimation method.

[6]Online Appendix (Table A5) shows sample means of covariates by whether the mother is identified. Individuals whose mother is not identified are more likely to be older, have higher income and have foreign background. Result do not qualitatively change if we use only individual's own covariates.

[7]From a design-based perspective, clustering may not be necessary as our treatment is assigned at the individual level (Abadie et al., 2022). However, in order to generalize our results to the whole population of young voters clustering accounts for municipality-level sampling variance as we observe only a subset of Finnish municipalities.

[8]As the number of treated individuals living together with control group individuals is small (5% of the control group individuals) even very large spillovers of over 100% would not affect our direct effect estimates at any meaningful decimal level. Thus, we do not examine potential spillovers from treatments groups to control groups, but focus on the intra-household transmission of treatment effects from our target sample (voters aged 18 to 29 years) to other eligible voters.

[9]This includes all household in which the combined number of control and treatment group individuals exceeds one. There are 861 households where there were two potentially treated individuals, 34 households with three potentially treated individuals and four households with four or more potentially treated individuals. Thus, in total, we drop 5.2% of the households from the analysis sample due to them having more than one potentially treated individual.

effects includes all individuals living within the same household in the end of year 2020 (as this is the most recent data point available to us) with an individual who received an SMS reminder and the control group consists of all individuals who were cohabiting with a young voter who was part of the control group.[10] On average, there are 1.52 voting aged individuals in addition to the SMS receiver (or control group member) in these households. Online Appendix (Table A8) shows that treatment and control groups used in the spillover analysis are balanced in terms of the covariates. We estimate the same set of models for the spillover sample as we do for the direct effects sample.

5.3 Effect heterogeneity analysis

The estimation of direct and spillover effects enables us to assess the effects of SMS reminders on turnout at large. However, these effects may not be evenly distributed in the electorate and may either exacerbate or ameliorate existing disparities in political participation. Building upon the work by Arceneaux and Nickerson (2009) and Enos et al. (2014), we analyze the effect of text message-based mobilization on the composition of the electorate. Our estimation procedure involves the following steps. First, we predict a propensity to vote for every individual using the available administrative data and the following logistic regression model:

$$Pr(Y_i = 1|\boldsymbol{X}_i) = \frac{exp(\boldsymbol{X}_b)}{1 + exp(\boldsymbol{X}_b)},$$

where $Pr(Y_i = 1|\boldsymbol{X}_i)$ is the predicted probability of voting based on individuals' gender, age, logarithm of (mother's) taxable income, ethnicity, education, SES background, eligibility to vote for the first time and municipality fixed effects. It is noteworthy that we are able to estimate these individual propensities to vote using a much richer set of personal information than what has been available in previous studies.

To estimate individual voting propensities in the absence of treatment, we conduct the propensity score estimation in a sample that is restricted to individuals assigned to the control group. The random assignment of individuals into the treatment and control groups guarantees that the propensity estimates in the control group are equally representative of the treatment group. Consequently, we compute for every individual in the sample their predicted probability to vote in the Finnish 2022 county elections in the absence of the SMS mobilization campaign. Second, we group the voting propensities by 25th,

[10]In the Online Appendix (Table A12) we report results from an alternative specification where we include all households having at least one potentially treated individual, not dropping households where this number is larger than one, and using fixed effects for the number of potentially treated individuals. The estimated coefficient of 1.4 p.p. is consistent with our main specification.

25-75th, and top 25th percentiles.[11] This grouping is done to detect possible non-linear effects by voting propensity (Arceneaux and Nickerson, 2009; Fowler, 2015). Splitting the sample into three groups is a more flexible approach compared to imposing a functional form for voting propensity by adding it into an OLS specification, while it retains statistical power for doing group comparisons compared to finer groupings. Finally, we estimate the effect of receiving an SMS reminder in these groups using the linear probability model to test whether the treatment systematically interacts with the existing disparities between high-propensity voters, marginal voters, and under-represented low-propensity voters. Even though this analysis is not pre-registered, it follows exactly the same procedure as in Hirvonen et al. (2023) in regards of estimating, predicting the voting propensities, and constructing the different groups.[12]

We note that the estimation of voting propensities through logistic regression may pose a risk of overfitting the data by fitting random variation and using outlier observations in demographic variables that could lead to biased comparison of treatment heterogeneities between high-propensity voters and under-represented voters. To address this concern, we complement the initial analysis by estimating voting probabilities through the Electic Net (Zou and Hastie, 2005; Hastie et al., 2015). The Elastic Net chooses an optimal combination of predictors using two penalty terms: one from LASSO (based on absolute value of the estimated coefficient, enabling elimination of predictors) and one from ridge regression methods (based on the square of the estimated coefficient, not enabling elimination of predictors). Thus, the Elastic Net overcomes, first, the tendency of LASSO to select only one predictor among highly correlated covariates. Second, the method allows dropping out predictors, which is not done by ridge regression alone. The procedure employs sample folding to separate the choice of parameters for penalty terms and fitting the model. Taken together, the Elastic Net trades bias for less variance by using penalty terms, reducing the risk of over-fitting the data.[13]

Moreover, we examine potential treatment effect heterogeneity also by using (pre-registered) subsamples that are created simply by using a single observed characteristic at a time. The pre-analysis plan registered at the American Economic Association Registry for RCTs mentions age, geographical area, previous voting history, education and income as potential grouping variables for heterogeneous

[11] This grouping splits the sample into half between the marginal group, where we would theoretically expect the largest effect, and the others. To study the robustness of our results based on this grouping, we use an alternative grouping that splits the sample into three equally sized group. The results using this alternative grouping are reported in the Online Appendix (Tables A3 and A4).

[12] Online Appendix (Tables A6, A7 and A9) provides means of used covariates by voting propensity group for the main and spillover samples.

[13] We report in the Online Appendix (Figure A1) the Receiving Operating Characteristic (ROC) curves for in-sample and out-of-sample predictions using the logit and elastic net models. We find that the Logit model is slightly better in terms of Area Under Curve (AUC) for in-sample prediction, whereas the elastic net model has higher AUC for the out-of-sample prediction. Online Appendix (Tables A6 and A7) shows the covariates by voting propensity groups for the logit and the elastic net models. The latter has a steeper gradient in terms of gender and educational background.

treatment effects. However, we did not present any specific hypotheses about the direction or magnitudes of potential effects. Lastly, we apply the honest causal forest approach by Wager and Athey (2018) to explore the heterogeneity of treatment effect using a multi-step procedure to avoid over-fitting the data. This data-driven method reduces the freedom of the researcher in terms of model specification.

6 Results

6.1 Direct effects

We begin by estimating the effect of SMS reminders on turnout at large and report the average treatment effect (ATE) in Table 3. We observe that receiving an SMS reminder leads to a 0.9 percentage point (p.p.) increase in turnout. This effect is statistically significant at the conventional 1% significance level. As expected, the ATE estimate remains stable around 0.9 p.p. after progressively adding demographic control variables. To put the effect size into perspective, we note that turnout in the control group is 30.8 percent. Thus, the effect size of 0.9 p.p. equals around 3% increase compared to the turnout in the control group. Moreover, we observe that receiving an SMS reminder bridges the gap between the 18-29-year-old voters and all other voters with an average turnout of 47.0% by 5.6%. Analogously, an SMS reminder bridges the gap between the 18-29-year-old voters and the 30-39-year-old voters with an average turnout of 36.6% by 16% – if the spillovers are not accounted for. Overall, our estimates are largely consistent with the findings from existing studies that have examined the effectiveness of text message reminders in the US and Nordic countries.

Next, we estimate direct treatment effects across the different treatment arms. Table 4 shows point estimates by treatment using the same control variables as in Column (3) in Table 3. We find that the treatment effect for the Neutral treatment is 1.6 p.p. and statistically significant at 1% significance level. This effect size is almost twice as large as for the Expressive treatment (0.9 p.p). However, the difference between the two estimates is not statistically significant at conventional significance levels. Moreover, we find that the point estimate for the Neutral treatment is eight times larger than the point estimate for the Instrumental treatment (0.2 p.p.). The difference between these two coefficients is statistically significant at 5% significance level. Overall, these observations suggest that the most simplified message

Table 3: Average Treatment Effect

	Voted			
	(1)	(2)	(3)	(4)
Treatment (pooled)	0.009***	0.009**	0.009***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
Municipality FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.307	0.308	0.308	0.308
Observations	50.140	49.679	49.679	49.679

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. This table follows a pre-analysis plan.

not appealing to any particular motivation to vote may have been the most effective at getting the young voters to turn out their vote.[14]

6.2 Spillover effects

In this section, we conduct similar analyses as in the previous section but apply the estimation procedure to measure within household spillover effects on untreated individuals. Table 5 shows that the ATE for the intra-household spillovers is around 1.4 p.p., suggesting that over 100 percent of the direct treatment effect spilled over to untreated household members. Alternatively, the relative size of the spillover effect can be put into perspective by comparing the size of the spillover estimate with the size of the direct effect estimate using a sample which includes only individuals residing in the spillover sample households. Using this sample, we find that the direct effect estimate is 0.6 p.p. (Online Appendix Table A11), suggesting that the relative size of the spillover effect is even larger in this sample. Finally, the spillover effect size

[14]To address a concern that the observed differences between the messages would be due to different opening rates caused by non-identical opening sentences of the SMS notifications, we conduct a supplementary analysis that exploits variation in the opening sentences between the advance period and election day messages (the first sentence of our SMS reminders is identical between the treatments in the election day messages, but there are differences between the first sentences in the advance voting period messages). Online Appendix (Table A14) shows that the differences between the treatment arms are similar in the advance period and election day voting data. Moreover, we revisit the differences between the messages in the Online Appendix (Table A10) by estimating spillover effects by treatment arm and find that the Neutral treatment leads to largest spillover effects.

Table 4: Different Treatments

	Voted			
	Pooled	Neutral	Expressive	Instrumental
	(1)	(2)	(3)	(4)
Treatment	0.009***	0.016***	0.009*	0.002
	(0.003)	(0.005)	(0.005)	(0.004)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.308	0.308	0.308	0.308
Observations	49.679	29.799	29.806	29.832
Differences		Neutral - Expressive	Expressive - Instrumental	Instrumental - Neutral
		0.007	0.007	-0.015**
		(0.007)	(0.006)	(0.007)

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. This table follows a pre-analysis plan.

of 1.4 p.p. equals around 2.8% increase compared to the baseline turnout of 49.6% in the control group (Table 5).

Table 6 divides the spillover sample by age and strength of social ties between the household members. We find a spillover effect of 1.8 p.p. (statistically significant at 10% error level) for current household members who are aged 50 years or above and a spillover effect of 1.0 p.p. for individuals who are aged below 50 years, suggesting that the spillover effects are more pronounced among older individuals (parents) living in the same household with the young voters.[15] Moreover, we test in Table 6 whether the spillovers extend to social ties beyond the current cohabitants. In Columns (3) and (4), we divide the sample into individuals which lived together with (exactly one) treated or control group individual in year 2014 but no more in year 2020. The estimated effect sizes, -0.5 p.p. for individuals aged 50 and above and 0.8 p.p. for individuals aged 49 or less, suggest that cohabitation is a relevant mechanism for the transmission of spillover effects within the social networks.[16]

The observed spillover effects lead to two important implications. First, in the presence of sizable spillover effects, impact evaluation analyses not able to detect spillovers among social ties may lead to a substantial underestimation of the net causal effect. Second, spillovers from the target populations (e.g., young voters) to other population groups (e.g., older voters) could mean that the gap in turnout between

[15] We do not observe in our data the exact type of relationship (biological vs. non-biological relationship) between the young voters and older individuals living in the same household and assume here that the individuals aged 50 and above are mainly parents living in the same household with the young voters.

[16] The observation about the importance of cohabiting in the transmission of spillover effects is consistent with an observation that there is a substantial and robust increase in turnout after moving together (Dahlgaard et al., 2022).

Table 5: Spillovers - Average Treatment Effect

	Voted			
	(1)	(2)	(3)	(4)
Treated in HH	0.014***	0.014***	0.013***	0.011***
	(0.006)	(0.005)	(0.006)	(0.005)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
Municipality FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.494	0.496	0.496	0.496
Observations	37.207	36.876	36.876	36.876

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. This table follows a pre-analysis plan.

the targeted population group and the other population groups does not shrink as much as suggested by simplistic comparisons based on estimated direct treatment effects – in principle the participation gap can even widen. At the same time, interventions with large spillovers may influence social inequalities in voting within the spillover group, raising new and less studied questions about the total effect of GOTV mobilization on the composition of the electorate.

6.3 Heterogeneous effects by voting propensities

To assess the impact of our intervention on inequality in voting, we estimate in this section heterogeneous treatments effects by voting propensity. Table 7 (Panel A) presents direct treatment effects for voters divided into three voting propensity groups - Low Propensity Voters, Marginal Voters and High Propensity Voters - based on a logit model.[17] Table 7 (Panel B) reiterates the same analysis for within household spillover estimates. We find that the direct effect estimate for the low propensity voters is 2.0 p.p.. The direct effect for the marginal voters is 1.2 p.p.. The former coefficient is statistically significant at 1% level, while the latter coefficient is statistically significant at 5% level. The point estimate for the high propensity voters is -0.8 p.p., albeit not statistically significantly different from zero. The estimates of the first two voting propensity groups are significantly different from the high propensity voters' estimate at

[17]Online Appendix (Table A3 and Table A4) shows results by three equal percentile splits. The results from these estimations are not qualitatively different.

Table 6: Spillover Heterogeneity by Age and Household Membership

	Voted			
	Current household member		Former household member	
	Over 49	Under 50	Over 49	Under 50
	(1)	(2)	(3)	(4)
Spillover Treatment	0.018	0.010	-0.005	0.008
	(0.011)	(0.006)	(0.008)	(0.007)
Controls	✓	✓	✓	✓
Untreated $\bar{Y}$	0.620	0.428	0.624	0.353
Observations	13.054	23.822	16.050	21.577
Differences	0.008		-0.013	
	(0.013)		(0.010)	

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

1% significance level for the low propensity group and at 5% significance level for the marginal propensity group. Given that the baseline turnout rate for the low propensity voters is only around one half of the marginal voters' turnout and less than one third compared to the high propensity voters, the relative effect size for the low propensity voters is remarkably larger than for the two other groups. Overall, our intervention seems to have reduced existing social inequalities in voting among the young voters - or at least it did not exacerbate existing inequalities in voting.

Table 7 (Panel B) shows heterogeneous treatment effects by voting propensities for the spillover sample after we estimate a voting propensity for every individual living in a same household together with a treated or untreated young voter and assign these cohabitants into three groups based on their predicted probability to vote. We observe that the low propensity group has a point estimate of 2.1 p.p., whereas the marginal voters group has a point estimate of 1.6 p.p. The estimate for high propensity voters is -0.6 p.p.. In the same manner as the direct treatment effects, the spillover effects seem to reduce turnout inequality among the unreated individuals.

The last panel (Panel C) of Table 7 shows spillover effects by the voting propensity of young voters potentially receiving an SMS reminder living in these households.[18] The results are consistent with the results from Panels A and B, and show that the spillover effects mainly originate from low propensity and marginal voters groups (households) as the point estimates are 1.3 p.p. and 1.6 p.p. for low propensity

[18]We cross-tabulate in the Online Appendix (Table A15) voting propensity groups from Panel B and Panel C of Table 7 and find it to be unlikely that spillover sample individuals with high voting propensity have potentially treated young household members from the low propensity group and vice versa. The young voters potentially receiving an SMS reminder include treatment and control group members.

Table 7: Heterogeneity by Vote Propensity

	Voted			
	All	Low Propensity {Bottom 25%}	Marginal Voters {25-75%}	High Propensity {Top 25%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.020***	0.012**	-0.008
	(0.003)	(0.007)	(0.005)	(0.008)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.309	0.151	0.299	0.485
Observations	49.458	12.363	24.727	12.368
Differences		Marginal - Low	Marginal - High	High - Low
		-0.008	0.020**	-0.028***
		(0.008)	(0.009)	(0.010)
Panel B: Spillover Effects by HH Members' Voting Propensity				
Treated in HH	0.013**	0.021**	0.016**	-0.006
	(0.006)	(0.010)	(0.008)	(0.011)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.497	0.242	0.495	0.761
Observations	36.723	9.180	18.362	9.181
Differences		Marginal - Low	Marginal - High	High - Low
		-0.005	0.022*	-0.027*
		(0.013)	(0.013)	(0.015)
Panel C: Spillover Effects by Targeted (Young) Voters' Voting Propensity				
Treated in HH	0.013**	0.013	0.016**	0.006
	(0.005)	(0.013)	(0.008)	(0.012)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.497	0.358	0.500	0.630
Observations	36.437	9.089	18.242	9.106
		Marginal - Low	Marginal - High	High - Low
Differences		0.003	0.010	-0.007
		(0.015)	(0.014)	(0.018)

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

and marginal voters groups, respectively. The estimated effect for the high propensity group is 0.6 p.p and not statistically different from zero at the conventional error levels. We report in the Online Appendix (Table A16) spillover effects for the combinations of groups as defined in Panel B and Panel C of Table 7. Although the sample sizes of these subgroups are smaller and the estimates less precise, the spillover effects seem to be most prevalent in households in which combinations of young low propensity or marginal voter reside together with older low propensity or marginal voters.

To alleviate the concern of over-fitting the data while estimating the predicted probabilities to vote, we reproduce the analysis reported in Table 7 using predictions estimated by Elastic Net (Zou and Hastie, 2005; Hastie et al., 2015). Table 8 shows results using this alternative estimation procedure. We observe in Table 8 (Panel A) that the group of Marginal Voters now has the highest point estimate of 1.5 p.p., which is statistically significantly different from zero at 1% significance level. This group of marginal voters is followed by the low propensity voters with an estimate of 0.6 p.p. and the high propensity voters with an estimate of -0.2 p.p.. The T-test for difference between marginal propensity voters and high propensity voters is statistically significant at 5% significance level. In the spillover sample (Panel B), marginal voters have the highest point estimate of 1.5 p.p., which is statistically different from zero at 10% significance level. Estimated coefficients for the low propensity and the high propensity groups are 1.3 p.p. and 0.9 p.p., respectively. The T-tests for differences between these three groups do not yield any statistically significant p-values. In panel C, the treatment effect for the marginal voters group is 2.2 p.p. (statistically significant at 1% significance level). For low and high propensity voter groups we observe estimates that are close to zero not statistically significant. We interpret these results as evidence against the conjecture that the spillovers originating from our SMS-based GOTV mobilization campaign would have widened disparities in participation. These results complement existing evidence related to the compositional effects of GOTV interventions and highlight the importance of studying spillover effects to gain a better understanding of the total effect of GOTV mobilizations on the composition of the electorate.

6.4 Heterogeneous effects by various subsamples

The preceding sections provided important insights about the effects of GOTV interventions on the composition of the electorate using data-driven estimation methods. These insights, however, may hold limited practical implications for the development of concrete strategies to ameliorate gaps in political participation. Our results so far suggest that text-message-based interventions may not only raise turnout but could also decrease the gap in voting. However, many organizations implementing GOTV interven-

Table 8: Heterogeneity by Vote Propensity - Elastic Net

	Voted			
	All	Low Propensity {Bottom 25%}	Marginal Voters {25-75%}	High Propensity {Top 25%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.006	0.015***	-0.002
	(0.003)	(0.006)	(0.005)	(0.007)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.308	0.161	0.294	0.481
Observations	49.679	12.361	24.806	12.512
Differences		Marginal - Low	Marginal - High	High - Low
		0.009	0.017**	-0.008***
		(0.008)	(0.008)	(0.009)
Panel B: Spillover Effects by HH Members' Voting Propensity				
Treated in HH	0.013**	0.013**	0.015*	0.009
	(0.006)	(0.011)	(0.009)	(0.012)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.496	0.247	0.491	0.753
Observations	36.876	9.219	18.438	9.219
Differences		Marginal - Low	Marginal - High	High - Low
		0.002	0.005	-0.004
		(0.014)	(0.015)	(0.016)
Panel C: Spillover Effects by Targeted (Young) Voters' Voting Propensity				
Treated in HH	0.013**	0.002	0.022***	0.001
	(0.006)	(0.012)	(0.007)	(0.011)
Controls	✓	✓	✓	✓
Untreated $\bar{Y}$	0.496	0.365	0.497	0.629
Observations	36,838	9,176	18,534	9,128
		Marginal - Low	Marginal - High	High - Low
Differences		0.020	0.021	-0.001
		(0.014)	(0.014)	(0.017)

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

tions with an aim to reduce demographic inequalities in voting are unlikely to have access to very detailed individual-level data before implementing their interventions. Moreover, the use of detailed individual-level data in data driven campaigning may pose difficult questions about the effect of campaigning on many established democratic principles and fair political processes (Kefford et al., 2023). Therefore, we also analyze effect heterogeneity using univariate sample splits.

Table 9 divides the sample by educational background, ethnicity, voting in 2021 municipality elections and type of residential area (urban vs. rural) each separately. By comparing Columns (1) and (2) in Table 9, we observe that the point estimates for the direct effects (Panel A) are identical for the two education groups and slightly higher for the spillover effects (Panel B) for individuals whose mother has a high school degree than for individuals whose mother did not finish high school. However, these estimates are not statistically significantly different from each other. Turning into ethnicity, we observe that individuals born in Finland to Finnish parents have positive point estimates for the direct effects (Panel A) and for the spillover effects (Panel B), whereas immigrants have a negative direct effect estimate (-0.9 p.p.) and a negative spillover estimate (-1.7 p.p.). However, the sample size for individuals with an immigration status is small and the observed negative coefficients are not statistically different from zero. For spillover effects, the coefficient for the difference between native and non-native individuals is statistically significant at 10% significance level. Overall, these observations suggest that the intervention could have widened the turnout gap between the immigrants and the natives. However, our reminders were sent in Finnish and Swedish, the two official languages in Finland, while all individuals aged 18 and above with a permanent residence in Finland are eligible to vote in the county elections. Thus, the eligibility did not depend on the citizenship and associated language requirements, which may have contributed to the widening participation gap between the immigrants and the natives.

Columns (5) and (6) of Table 9 presents results by voting in 2021 municipality elections and urbanity of the municipality. Panel A shows estimates for direct effects. Point estimate for individuals who voted in the 2021 elections is 2.8 p.p. and statistically significant at 1% significance level. The point estimate for those who did not vote in 2021 elections is 0.6 p.p. and not statistically different from zero. The difference of coefficients is statistically significant at 1% significance level. Panel B shows the results for spillover estimation. We find that those who voted in 2021 have a higher spillover effect with a point estimate of 2.0 p.p. compared to a coefficient of 0.8 p.p. for those who did not vote in 2021. This provides some evidence for the experiment having a widening effect on the participation gap.

Columns (7) and (8) in Table 9 split the sample into individuals living in rural and urban municipalities. In Panel A, the treatment estimate for residents in urban municipalities is higher (1.2 p.p.)

Table 9: Heterogeneous Effects by Subsamples

	Educational background		Ethnicity		Voting in 2021		Urbanity	
	High School	No High S.	Native	Non-native	Voted	Not voted	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Direct Effects								
Treated	0.009*	0.009	0.010***	-0.009	0.028***	0.006	0.005	0.012***
	(0.005)	(0.006)	(0.003)	(0.013)	(0.007)	(0.004)	(0.014)	(0.004)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Control group $\bar{Y}$	0.405	0.230	0.316	0.120	0.593	0.114	0.313	0.306
Observations	22,020	27,659	47.696	1.983	17.643	27.800	5.335	38,791
Differences	0.001		0.019		0.022***		-0.007	
	(0.007)		(0.014)		(0.008)		(0.014)	
Panel B: Spillover Effects								
Treated in HH	0.015**	0.012	0.014**	-0.017	0.020***	0.008	0.034*	0.012**
	(0.007)	(0.008)	(0.006)	(0.018)	(0.007)	(0.007)	(0.018)	(0.006)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Control group $\bar{Y}$	0.620	0.410	0.510	0.175	0.745	0.164	0.531	0.496
Observations	15,065	21,811	353.24	1.552	20.646	15.170	5.599	29.418
Differences	0.004		0.032*		0.012		0.023	
	(0.011)		(0.019)		(0.010)		(0.019)	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion), and an indicator variable whether individual was eligible to vote for the first time. This table follows a pre-analysis plan.

compared to young voters residing in rural municipalities (0.5 p.p.). For the case of spillovers in Panel B, we find an opposite effect as individuals living in rural municipalities have a higher point estimate (3.4 p.p.) compared to individuals living in urban municipalities (1.2 p.p.). However, the differences between the estimated coefficients both in the direct effect sample and in the spillover sample are not statistically significant.

Additionally, we employ a more data-driven machine learning approach for the estimation of potential heterogeneous treatment effects. Online Appendix (Figure A2 and Figure A3) presents results from the honest causal forest method: both for direct and spillover samples there are no statistically significant differences between CATE ranking groups indicating that results are too imprecise to draw any conclusions about heterogeneous treatment effects.

7 Discussion

This paper employs four different estimation methods to detect heterogeneous treatment effects. First, we split the sample in three vote propensity groups using a logit model. Second, we reproduce the same procedure using an elastic net model. We find that predicting the voting propensities using a logit model led to the highest point estimate for the low propensity group. In addition, we find that predicting the voting propensities using the elastic net method led to the highest point estimate for the marginal propensity group. Both of these two complementary methods yielded for the high propensity voters estimates that were close to zero and not statistically significant. The differences arise from using municipality fixed effects, which improve voting prediction.[19] Around 83% of the sample remains in the same vote propensity group regardless of which of the two prediction models is used. The difference between the prediction models is due to the LASSO regularization part of the elastic net dropping some of the municipality fixed effects by setting the corresponding fixed effect coefficients equal to zero. Without the municipality fixed effects, there are practically no differences between these two methods as over 99.99% of observations are assigned to the same vote propensity groups.

A mere comparison of the point estimates might not give the full picture about relative effects as the number of individuals differ by vote propensity groups, and direct effect and spillover samples are not of equal size. Thus, we define the number of activated voters as: $\beta_1(Treatment) \times N_{treated\_individuals}$ in each group. Online Appendix (Table A18) presents the number of activated voters by vote propensity groups for direct effect and spillover effect samples, and by age for the spillover sample using the logit estimates.

[19]Leaving out municipality fixed effects results in around 17% smaller adjusted-R^2 in a linear prediction model.

When we account only for the direct effect, the experiment resulted in an increase of around 150 voters in the low propensity group and 177 voters in the group of marginal voters. As the heterogeneity estimates are remarkably similar to those from direct effect sample with logit vote propensity groups and elastic net vote propensity groups, it means that voting inequality was not just reduced by the direct effect but also among individuals subject to the spillovers.[20] In terms of activated voters, the experiment resulted in 116 more votes from the low propensity and 177 votes from the marginal voters spillover groups. Overall, accounting for the spillover effects more than doubles the total number of activated voters from 268 to 557 voters. Additionally, spillovers have an implication on how the voting gap between the target population of youth voters and older voters evolved. When also the within household spillovers are taken into account, the experiment activated 90 more under 30-year-old voters compared to over 29-year-old voters. This is a sizable relative reduction from the net of 268 activated youth voters when spillovers are ignored. However, it should be noted that the experiment differs from a policy where every youth voter is reached by SMS, as for that policy there is no possibility for spillovers to under 30-year-old voters who did not receive a voting reminder. Additionally, we are not able to determine the precise effect for the voting gap between youth and old voters in relative terms, as the spillover group is not a random sample from the general population of over 29-year-old eligible voters.

Third, we utilize a machine learning method called Honest Causal Forest to detect heterogeneous treatment effects. While this algorithm has nice properties in terms of reducing freedom of the researcher when specifying a model, in our case, we are unable to draw conclusions from the results as estimates obtained from the model are very imprecise. This is unsurprising as even though the number of CATE groups is the same as the number of vote propensity groups, the effective number of observations is lower as only part of the sample is used for the estimation due to sample splitting and folding.

Lastly, we use sub-group analysis where we split the sample by a set of covariates. At first glance, the higher estimate for the group that had voted in 2021 municipality elections could be seen as contradictory to our vote propensity group results coming from low and mid propensity voters. How can we reconcile these contrasting conclusions about the effects of our intervention on inequality in voting? In Online Appendix (Table A17), we split the three voting propensity groups by the past voting variable and provide effectiveness estimates for all the resulting six groups. The highest point estimates of 6.2 p.p. and 3.0 p.p. (both statistically different from zero at 1% error level) are, respectively, for the low propensity and marginal voters groups who also voted in 2021. The control group mean is remarkably higher for these

[20] It should be noted that these spillover results do not necessarily generalize to the entire population aged 29 or above, nor to a policy setting where everyone, regardless of age, would receive a reminder.

groups compared to the individuals belonging to the same vote propensity group but who did not vote in 2021. Also these two latter groups have estimates which are positive and statistically significant, albeit smaller. However, for high propensity voters we do not find statistically significant estimates for either those who voted in 2021 nor did not vote in 2021. These results indicate that among the low and mid propensity groups the treatment effect is largely coming from individuals who tend to vote. Therefore, the effects are coming from citizens who clearly have intention to vote, but whose characteristics would indicate that they are unlikely to vote. This is consistent with the Noticeable Reminder Theory (Dale and Strauss, 2009) that the effect comes from citizens who have an intention to vote, but may fail to do so, because of lack of attention for example, in which case only a simple nudge is enough to remind them of their intention to vote. It is also consistent with the The Receive-Accept-Sample Theory (Zaller, 1992) as citizens belonging to this social group may not receive too many messages, and thus, accepting the message is particularly likely for the voters who have high individual interest in voting, but live in low interest environment. Taken together, this combination of high interest but low exposure could make nudging more efficient.

8 Conclusions

This paper presents new evidence about the spillover and compositional effects of GOTV interventions. We contribute new evidence to the discussion on inequalities in political participation. Moreover, using an RCT design and data-driven estimation techniques, we provide new insights to the debate on the transmission of spillover effects in voting. We obtain several empirical findings. First, we find that receiving a text message reminder before the Finnish country elections in 2022 mobilized low-propensity and marginal voters and reduced, or at least did not exacerbate, existing social inequalities in voting within our target sample, 18 to 29-year-old voters. Second, the effects are dominantly coming from citizens belonging to low vote propensity social strata, but who have high personal intention to vote. Third, we document remarkably large spillover effects in voting behavior, suggesting that the behavior of adult children with voting rights may influence especially their parents' turnout decisions. Fourth, the spillover effects reduced inequalities among the group affected by the spillovers. Fifth, we find suggestive evidence that the most simplified phrasing of messages merely reminding recipients about the approaching elections was more effective than the messages appealing to expressive and instrumental motivations to vote. One potential explanation for this finding is aversion to paternalism among the citizens.

Our paper is complimentary to studies that have previously examined the compositional effects of GOTV interventions and transmission of voting behavior in social networks. We document that implicitly assuming zero or little spillovers among social ties may underestimate the true effectiveness of voter mobilization interventions. Moreover, our results show that the average estimates of spillover effects may mask considerable heterogeneity in the peer-to-peer transmission of voting behavior, suggesting that studies that do not account for spillover effects may misreport the compositional effects of GOTV interventions. Our observation that the spillovers occur mainly within households and run from young adults to their parents speaks to the potential trickle-up effects in voting between the generations and may help understanding the role of trickle-up effects in political socialization (Jennings and Niemi, 1981; Linimon and Joslyn, 2002; Dahlgaard, 2018).

Our results suggest that the previously observed compositional effects of GOTV interventions that have widened the disparities in participation do not readily generalize to our context. In fact, we observe that SMS reminders are effective at mobilizing low-propensity voters and their household members. This observation is consistent with the existing Nordic studies reporting that GOTV interventions may mostly mobilize low-propensity voters (Bhatti et al., 2018; Bergh et al., 2020; Bergh and Christensen, 2024). However, unlike the previous European experiments, our experiment is conducted in a relatively low-salience elections, suggesting that the discrepancy between the findings from the US-based experiments (Enos et al., 2014) and European experiments is not explained solely by the differences in the average turnout in the electorate. One potential explanation for the difference is that Nordic countries have high levels of trust in government and public institutions. This mechanism should be evaluated in further studies. Overall, our results hold promise that impersonal but inclusive means of communication, like text messages, may not only successfully raise aggregate voter turnout but also encourage less likely voters to turn out their vote.

Our results stem from an electoral context that is characterized by notably high trust in public administration. Moreover, it is worthy of highlighting that our intervention was developed in collaboration with the Finnish electoral authority and we sent the messages on behalf of the Ministry of Justice. Thus, our project can be classified as a state-led intervention that could produce different results in other countries and contexts with different levels of trust in public administration. Relatedly, there is accumulating evidence that text messaging by public authorities may lead to larger effects in voter registration than messages by advocacy organizations (Cheng-Matsuno et al., 2023). Overall, it is important to recognize that the results from our state-led intervention conduced in a Nordic country may not readily generalize to institutional contexts with different demographic composition of voters and mobilization campaigns organized by non-governmental organizations.

More generally, our paper advances the literature that examines how different sub-populations respond to a given treatment and assesses the potential of enhancing the effectiveness of behavioral interventions through selective targeting of existing interventions. We believe that the blend of RCT designs, comprehensive individual-level administrative datasets and suitable high-resolution predictive methods constitute a promising approach to enhance the effectiveness of interventions aiming to motivate behavioral change. Moreover, we observe that the combination of prediction models and univariate sample splits may help form a more detailed picture about the activated voters.

Our results raise new directions for future research. Our results hold promise that text message-based interventions may not only successfully raise turnout but also reduce inequalities in voting. A natural step towards better understanding the promises and limits of GOTV interventions as a tool to ameliorate demographic gaps in political participation is to study hard-to-reach populations who may be beyond the reach of conventional GOTV interventions, but are accessible through their mobile phones. Attempts to address the minuscule political participation in certain hard-to-reach populations, like young immigrants, may also substantially benefit from the tailoring of treatment designs (e.g., use of their native language) to these specific subgroups.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge, "When Should You Adjust Standard Errors for Clustering?," *The Quarterly Journal of Economics*, 10 2022, *138* (1), 1–35.

Arceneaux, Kevin and David W Nickerson, "Who is mobilized to vote? A re-analysis of 11 field experiments," *American Journal of Political Science*, 2009, *53* (1), 1–16.

Bergh, Johannes and Dag Arne Christensen, "Getting out the vote in different electoral contexts: The effect of impersonal voter mobilization techniques in middle and high salience Norwegian elections," *Journal of Elections, Public Opinion and Parties*, 2024, *34* (1), 79–95.

_ , _ , and Richard E Matland, "Inviting immigrants in: Field experiments in voter mobilization among immigrants in Norway," *Electoral Studies*, 2020, *66*, 102160.

_ , _ , and _ , "When is a reminder enough? Text message voter mobilization in a European context," *Political Behavior*, 2021, *43* (3), 1091–1111.

Bhatti, Yosef, Jens Olav Dahlgaard, Jonas Hedegaard Hansen, and Kasper M Hansen, "How voter mobilization from short text messages travels within households and families: Evidence from two nationwide field experiments," *Electoral Studies*, 2017, *50*, 39–49.

_ , _ , _ , and _ , "Moving the campaign from the front door to the front pocket: Field experimental evidence on the effect of phrasing and timing of text messages on voter turnout," *Journal of Elections, Public Opinion and Parties*, 2017, *27* (3), 291–310.

_ , _ , _ , and _ , "Can governments use Get Out The Vote letters to solve Europe's turnout crisis? Evidence from a field experiment," *West European Politics*, 2018, *41* (1), 240–260.

Blais, Andre, Robert Young, and Miriam Lapp, "The calculus of voting: An empirical test," *European Journal of Political Research*, 2000, *37* (2), 181–201.

Brennan, Geoffrey and Michael Brooks, "Expressive voting," in "The Elgar Companion to Public Choice, Second Edition," Edward Elgar Publishing, 2013.

Cheng-Matsuno, Vanessa, Florian Foos, Peter John, and Asli Unan, "Do text messages increase voter registration? Evidence from RCTs with a local authority and an advocacy organisation in the UK," *Electoral Studies*, 2023, *81*, 102572.

Dahlgaard, Jens Olav, “Trickle-up political socialization: The causal effect on turnout of parenting a newly enfranchised voter,” *American Political Science Review*, 2018, *112* (3), 698–705.

_ **, Yosef Bhatti, Jonas Hedegaard Hansen, and Kasper M Hansen**, “Living together, voting together: Voters moving in together before an election have higher turnout,” *British Journal of Political Science*, 2022, *52* (2), 631–648.

Dale, Allison and Aaron Strauss, “Don’t forget to vote: Text message reminders as a mobilization tool,” *American Journal of Political Science*, 2009, *53* (4), 787–804.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper, 1957.

Duflo, Esther, Rachel Glennerster, and Michael Kremer, “Using randomization in development economics research: A toolkit,” *Handbook of Development Economics*, 2007, *4*, 3895–3962.

Enos, Ryan D, Anthony Fowler, and Lynn Vavreck, “Increasing inequality: The effect of GOTV mobilization on the composition of the electorate,” *The Journal of Politics*, 2014, *76* (1), 273–288.

Foos, Florian and Eline A. de Rooij, “All in the Family: Partisan Disagreement and Electoral Mobilization in Intimate Networks—A Spillover Experiment,” *American Journal of Political Science*, 2017, *61* (2), 289–304.

Fowler, Anthony, “Electoral and policy consequences of voter turnout: Evidence from compulsory voting in Australia,” *Quarterly Journal of Political Science*, 2013, *8* (2), 159–182.

_ , “Regular voters, marginal voters and the electoral effects of turnout,” *Political Science Research and Methods*, 2015, *3* (2), 205–219.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer, “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment,” *American Political Science Review*, 2008, *102* (1), 33–48.

Green, Donald P and Alan S Gerber, *Get out the vote: How to increase voter turnout*, Brookings Institution Press, 2019.

_ **, Mary C McGrath, and Peter M Aronow**, “Field experiments and the study of voter turnout,” *Journal of Elections, Public Opinion and Parties*, 2013, *23* (1), 27–48.

Harjunen, Oskari, Tuukka Saarimaa, and Janne Tukiainen, "Love Thy (Elected) Neighbor? Residential Segregation, Political Representation, and Local Public Goods," *The Journal of Politics*, 2023, *Forthcoming.*

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright, "Statistical learning with sparsity," *Monographs on statistics and applied probability*, 2015, *143*, 143.

Hirvonen, Salomo, Jerome Schafer, and Janne Tukiainen, "Policy Feedback and Civic Engagement: Evidence from the Finnish Basic Income Experiment," Discussion Paper 155, Aboa Centre for Economics 2023.

Jääskeläinen, Arto, *The Finnish Election System: Overwiew*, Oikeusministerio, 2020.

Jennings, M Kent and Richard G Niemi, *Generations and politics: A panel study of young adults and their parents*, Princeton University Press, 1981.

Kefford, Glenn, Katharine Dommett, Jessica Baldwin-Philippi, Sara Bannerman, Tom Dobber, Simon Kruschinski, Sanne Kruikemeier, and Erica Rzepecki, "Data-driven campaigning and democratic disruption: Evidence from six advanced democracies," *Party Politics*, 2023, *29* (3), 448–462.

Klofstad, Casey A., "Talk Leads to Recruitment: How Discussions about Politics and Current Events Increase Civic Participation," *Political Research Quarterly*, 2007, *60* (2), 180–191.

Lassen, David Dreyer and Daniel Mahler, "Free to choose or free to lose? Understanding individual attitudes toward paternalism," *Behavioural Public Policy*, 2023, *7* (3), 721–743.

Lim, Stephen S, Rachel L Updike, Alexander S Kaldjian, Ryan M Barber, Krycia Cowling, Hunter York, Joseph Friedman, R Xu, Joanna L Whisnant, Heather J Taylor et al., "Measuring human capital: A systematic analysis of 195 countries and territories, 1990–2016," *The Lancet*, 2018, *392* (10154), 1217–1234.

Linimon, Amy and Mark R Joslyn, "Trickle up political socialization: The impact of kids voting USA on voter turnout in Kansas," *State Politics & Policy Quarterly*, 2002, *2* (1), 24–36.

Lyytikäinen, Teemu and Janne Tukiainen, "Are voters rational?," *European Journal of Political Economy*, 2019, *59*, 230–242.

Malhotra, Neil, Melissa R Michelson, Todd Rogers, and Ali Adam Valenzuela, "Text messages as mobilization tools: The conditional effect of habitual voting and election salience," *American Politics Research*, 2011, *39* (4), 664–681.

Mann, Christopher B and Katherine Haenschen, "A meta-analysis of voter mobilization tactics by electoral salience," *Electoral Studies*, 2024, *87*, 102729.

Matland, Richard E. and Gregg R. Murray, "An Experimental Test of Mobilization Effects in a Latino Community," *Political Research Quarterly*, 2012, *65* (1), 192–205.

Mo, Cecilia Hyunjung, John B Holbein, and Elizabeth Mitchell Elder, "Civilian national service programs can powerfully increase youth voter turnout," *Proceedings of the National Academy of Sciences*, 2022, *119* (29), e2122996119.

Naess, Ole-Andreas Elvik, "Increasing turnout with a text message: Evidence from a large campaign from the government," *Journal of Elections, Public Opinion and Parties*, 2022, pp. 1–19.

Nickerson, David W, "Is voting contagious? Evidence from two field experiments," *American Political Science Review*, 2008, *102* (1), 49–57.

Sinclair, Betsy, Margaret McConnell, and Donald P Green, "Detecting spillover effects: Design and analysis of multilevel experiments," *American Journal of Political Science*, 2012, *56* (4), 1055–1069.

Thaler, Richard H and Cass R Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Penguin, 2009.

von Schoultz, Åsa and Kim Strandberg, *Political Behaviour in Contemporary Finland: Studies of Voting and Campaigning in a Candidate-Oriented Political System*, Routledge, 2024.

Wager, Stefan and Susan Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

Zaller, John R., *The Nature and Origins of Mass Opinion*, Cambridge University Press, 1992.

Zou, Hui and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, 2005, *67* (2), 301–320.

Zuckerman, Alan S., *The Social Logic of Politics: Personal Networks as Contexts for Political Behavior*, Temple University Press, 2005.

Online Appendix for

Who is mobilized to vote by short text messages?

Evidence from a nationwide field experiment with young voters

Contents

A Supplementary results

A.1 Supplementary Tables

A.1.1 Alternative Model Specifications and Voting Propensity Groups

Table A1 and Table A2 present Average Treatment Effect (ATE) results by using logit model instead of la linear specification used in the main results. Both direct and spillover ATE results are robust to the choice of the model. Table A3 and Table A4 show voting propensity heterogeneity models with equall size vote propensity groups. For the logit prediction model, the point estimate is highest for the Low Propensity group followed by the Marginal Voters group and the coefficient for High Propensity group is negative and not statistically different from zero. For the elastic net model, estimates are similar across the voting propensity groups.

Table A1: Average Treatment Effect - Logit Model

	Voted			
	(1)	(2)	(3)	(4)
Treatment (pooled)	0.041***	0.042**	0.045***	0.047***
	(0.015)	(0.016)	(0.016)	(0.017)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
Municipality FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.307	0.308	0.308	0.309
Observations	50.140	49.679	49.679	49.599

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time. This table follows a pre-analysis plan.

Table A2: ATE for spillovers - Logit Model

	Voted			
	(1)	(2)	(3)	(4)
Treatment (pooled)	0.054**	0.060**	0.060**	0.052**
	(0.025)	(0.024)	(0.026)	(0.026)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
Municipality FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.494	0.496	0.496	0.496
Observations	37.207	36.876	36.876	36.796

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time. This table follows a pre-analysis plan.

Table A3: Heterogeneity by Thirds of Vote Propensity

	Voted			
	All	Low Propensity {Bottom 33%}	Marginal Voters {33-67%}	High Propensity {Top 33%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.022***	0.010**	-0.006
	(0.003)	(0.007)	(0.005)	(0.007)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.309	0.166	0.298	0.483
Observations	49.458	16.321	16.814	16.323
Differences		Marginal - Low	Marginal - High	High - Low
		-0.012	0.016*	-0.027***
		(0.008)	(0.008)	(0.010)
Panel B: Spillover Effects				
Treated in HH	0.013**	0.025***	0.014**	-0.003
	(0.006)	(0.008)	(0.008)	(0.010)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.497	0.272	0.494	0.727
Observations	36.723	12.118	12.486	12.119
Differences		Marginal - Low	Marginal - High	High - Low
		-0.011	0.017	-0.028**
		(0.012)	(0.013)	(0.013)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

Table A4: Heterogeneity by Thirds of Vote Propensity - Elastic Net

	Voted			
	All	Low Propensity {Bottom 33%}	Marginal Voters {33-67%}	High Propensity {Top 33%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.009	0.009	0.009
	(0.003)	(0.007)	(0.006)	(0.006)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.308	0.176	0.294	0.451
Observations	49.679	16.266	16.718	16.695
Differences		Marginal - Low	Marginal - High	High - Low
		-0.000	0.001	0.000
		(0.009)	(0.008)	(0.009)
Panel B: Spillover Effects				
Treated in HH	0.013**	0.013	0.017**	0.007
	(0.006)	(0.009)	(0.008)	(0.010)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.496	0.277	0.494	0.720
Observations	36.876	12.169	12.537	12.170
Differences		Marginal - Low	Marginal - High	High - Low
		0.004	0.009	0.006
		(0.012)	(0.013)	(0.013)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

A.1.2 Covariates by Mother Identified and Vote Propensity Groups

Table A5 shows covariates of direct effect sample by whether individual's mother is identified from the data. Table A6 and Table A7 present covariates by vote propensity groups for the same sample with logit and elastic net prediction models respectively. There are notably more females among marginal and high propensity voters compared to the spillover sample from Table A9. At the same time, income is much better predictor of the voting propensity in the spillover sample compared to the full sample. This observation could be expected as the current income of young voters is less likely to be equally strong predictor of their lifetime income as for the older people.

Table A5: Covariates by Mother Identified

	Mother Known	Mother Not Known
Female	0.39	0.43
	(0.49)	(0.49)
Age	23.15	27.47
	(2.57)	(2.00)
Highschool degree	0.33	0.39
	(0.47)	(0.49)
Taxable Income	12410.49	21910.64
	(11599.83)	(13618.41)
Immigrant	0.03	0.07
	(0.16)	(0.25)
Observations	33.509	17.592

Notes: Standard deviation in parenthesis.

Table A6: Covariates by Voting Propensity - Logit

	Low Propensity	Marginal Voters	High Propensity
Female	0.14	0.40	0.69
	(0.35)	(0.49)	(0.46)
Age	24.92	24.54	24.47
	(3.03)	(3.10)	(3.30)
High school Background	0.05	0.41	0.91
	(0.21)	(0.49)	(0.29)
Taxable Income	18837.28	15423.02	13387.01
	(14680.37)	(12725.48)	(11706.14)
Immigrant Background	0.15	0.01	0.00
	(0.35)	(0.08)	(0.02)
Observations	12.476	25.046	12.487

Note: Standard deviation in parenthesis.

Table A7: Covariates by Voting Propensity - Elastic Net Logit

	Low Propensity	Marginal Voters	High Propensity
Female	0.06	0.42	0.72
	(0.25)	(0.49)	(0.45)
Age	24.89	24.51	24.58
	(3.03)	(3.12)	(3.30)
High school Background	0.01	0.39	0.97
	(0.10)	(0.49)	(0.18)
Taxable Income	19306.74	15185.56	13388.72
	(14947.32)	(12577.50)	(11550.46)
Immigrant Background	0.15	0.01	0.00
	(0.35)	(0.09)	(0.02)
Observations	12.768	25.357	12.673

Note: Standard deviation in parenthesis.

A.1.3 Spillover Sample and Additional Results

Table A8 shows that randomization works for control and treatment group of the spillover sample as they are balanced in characteristics. Table A9 presents the spillover sample covariates by logit prediction vote propensity groups. We can observe from Table A10 estimated spillover effects by treatment type. As for the direct treatment effects, the estimated spillover effect for the Neutral treatment is higher than for the two other treatments. The estimated effect size of 2.1 p.p. is statistically significant at 1% significance level.

Table A11 provides an alternative benchmark for the size of the spillover effect as it shows direct effect ATE estimates for (potentially treated) youth from the spillover sample households. The estimated coefficient is 0.6 p.p. (not statistically different from zero at the conventional error levels) meaning that the spillover effect is even larger in relative terms when compared to this sample. Table A12 shows spillover ATE estimates without restricting the spillover sample to households with exactly one potentially treated youth individual. Column (4) shows a model, where fixed effects is added for the number of potentially treated youth individuals in the household. Estimates are comparable to the main results. Lastly, Table A13 presents spillover heterogeneous effects by the age cut-off of 29 years.

Table A8: Spillover Sample Balance Tests

	(1) All	(2) Control	(3) Treated	(4) Difference
Gender	0.497	0.497	0.497	-0.000
	(0.500)	(0.500)	(0.500)	(0.005)
Age	43.468	43.443	43.484	-0.040
	(14.443)	(14.470)	(14.426)	(0.154)
Education Background	0.409	0.410	0.408	0.002
	(0.492)	(0.492)	(0.491)	(0.005)
Taxable Income	34089.863	34233.644	33994.990	238.654
	(24100.358)	(24365.424)	(23923.930)	(257.970)
Immigration Background	0.042	0.042	0.042	-0.001
	(0.201)	(0.200)	(0.201)	(0.002)
Observations	36.876	14.662	22.214	36.876

Note: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard deviation in parenthesis for columns (1) to (3) and standard error for column (4).

.

Table A9: Spillover Sample Covariates by Voting Propensity - Logit

	Low Propensity	Marginal Voters	High Propensity
Gender	0.45	0.48	0.57
	(0.50)	(0.50)	(0.49)
Age	30.45	44.41	54.81
	(10.04)	(13.86)	(6.92)
Education Background l	0.09	0.37	0.80
	(0.28)	(0.48)	(0.40)
Taxable Income	21557.95	31214.75	52178.04
	(15509.72)	(19080.58)	(29122.75)
Immigration Background	0.16	0.01	0.00
	(0.36)	(0.08)	(0.01)
Observations	9.180	18.362	9.181

Note: Standard deviation in parenthesis.

Table A10: Spillovers - Different Treatments

	Voted			
	Pooled	Neutral	Expressive	Instrumental
	(1)	(2)	(3)	(4)
Treated in HH	0.013**	0.021***	0.006	0.012
	(0.006)	(0.008)	(0.008)	(0.008)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.496	0.496	0.496	0.496
Observations	36.876	22.028	22.113	22.059
Differences		Neutral -	Expressive -	Instrumental
		Expressive	Instrumental	Neutral
		0.016	-0.006	-0.010
		(0.011)	(0.011)	(0.011)

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. This table follows a pre-analysis plan.

Table A11: Youth in Spillover Sample Households - Average Treatment Effect

	Voted			
	(1)	(2)	(3)	(4)
Treatments Pooled	0.006	0.006	0.006	0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
Municipality FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.320	0.321	0.321	0.321
Observations	28.564	28.302	28.302	28.302

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Sample consists of treatment and control group youth voters who are present in spillover sample households. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

Table A12: Spillover All Households - Average Treatment Effect

	Voted			
	(1)	(2)	(3)	(4)
Treatments Pooled	0.022***	0.018***	0.017***	0.014**
	(0.006)	(0.006)	(0.006)	(0.006)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
N Potentially Treated in HH FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.498	0.499	0.499	0.499
Observations	39.943	39.591	39.591	39.591

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Sample consists of all households which had potentially treated individuals. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

Table A13: Spillover Heterogeneity by Under 30

	Outcome: Voted	
	Age	
	Over 29	Under 30
	(1)	(2)
Spillover Treatment	0.016*	0.007
	(0.009)	(0.009)
Controls	Yes	Yes
Untreated $\bar{Y}$	0.578	0.342
Observations	24,100	12,776
Differences	0.010	
	(0.013)	

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. This table follows a pre-analysis plan.

A.1.4 Heterogenous Results by Message Type and Advance Voting vs. Election Day Voting

A valid concern is that opening rates of SMS would be affected by the first words in the first sentence of the messages visible in push-in notifications. For instance, the pre-advance voting period expressive and instrumental treatment message does not convey information about the election in the same manner in the first sentence of the message as the neutral treatment. However, the first sentence of the pre-election day messages is identical across the treatments. If the differences in the first sentence would be driving the heterogeneity of results by the treatment type, we would expect that these differences would hold for pre-advance voting period messages but not for pre-voting day messages. To test this we conduct an analysis where we estimate the effects separately for voting in advance and voting on election day by each treatment type. Table A14 shows that there are no statistically significant detectable differences between advance and election day voting within the treatment types, thus suggesting that the differences in estimated coefficients between treatments are not driven by differences in opening rates due to different beginning of the messages.

Table A14: Advance and Election Day Voting

Treatment:	Pooled (1)	"Neutral" (2)	"Expressive" (3)	"Informative" (4)
Outcome: Voted in Advance				
Treated	0.004*	0.009**	0.002	0.002
	(0.003)	(0.003)	(0.004)	(0.004)
Control group $\bar{Y}$	0.137	0.137	0.137	0.137
Outcome: Voted on Election Day				
Treated	0.005	0.008	0.007*	-0.000
	(0.003)	(0.006)	(0.004)	(0.004)
Control group $\bar{Y}$	0.171	0.171	0.171	0.171
Controls	Yes	Yes	Yes	Yes
Observations	49.679	29.799	29.806	29.832
Differences	-0.000	0.001	-0.004	0.002
	(0.004)	(0.007)	(0.005)	(0.006)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, educational background (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

A.1.5 Spillovers by Combinations of Voting Propensity Groups and Past Voting

Table A15 shows cross-tabulation of (logit) voting propensity groups for young voters potentially receiving an SMS reminder (both treatment and control group members included) residing in spillover sample households and for spillover sample individuals. We can observe that it is unlikely for spillover sample individuals with high voting propensity to reside with potentially treated young household members from the low propensity group and vice versa. Table A16 shows spillover effects by combinations of aforementioned vote propensity groupings. Although the sample sizes of these groups are fairly small and the estimates less precise, the spillover effects seem to be most prevalent in households in which combinations of young low propensity or marginal voter reside together with older low propensity or marginal voters.

Table A17 presents evidence of effect estimates when we split logit voting propensity groups by voted in 2021. All four group estimates for the low propensity and marginal voters are statistically significant. However the effect size is much higher for those individuals who voted in 2021 in for low propensity and marginal voters with estimates of 6.2 p.p. and 3 p.p (both statistically different from zero at 1% error level). For high propensity voters we don't find any statistically significant estimates .

Table A15: Combinations of Voting Propensity Groups

		Youth in HH		
		Low Propensity	Marginal Voter	High Propensity
Spillover HH Member	Low Propensity	0.354	0.502	0.144
	Marginal Voter	0.235	0.505	0.260
	High Propensity	0.085	0.488	0.426

Table A16: Spillovers by Combinations of Voting Propensity Groups

	Voted			
	All	Low Propensity {Bottom 25%}	Marginal Voters {25-75%}	High Propensity {Top 25%}
	(1)	(2)	(3)	(4)
Panel A: Low Propensity Youth in HH				
Treated in HH	0.012	0.007	0.030*	-0.099**
	(0.013)	(0.013)	(0.018)	(0.038)
Controls	Yes	Yes	Yes	Yes
Control group $\bar{Y}$	0.360	0.192	0.423	0.748
Observations	9.043	3.447	4.827	769
		Marginal - Low	Marginal - High	High - Low
Differences		0.023	0.129***	-0.106***
		(0.022)	(0.042)	(0.040)
Panel B: Marginal Youth in HH				
Treated in HH	0.015**	0.027*	0.011	0.013
	(0.008)	(0.014)	(0.010)	(0.016)
Controls	Yes	Yes	Yes	Yes
Control group $\bar{Y}$	0.501	0.256	0.504	0.731
Observations	18.184	4.318	9.412	4.454
		Marginal - Low	Marginal - High	High - Low
Differences		-0.016	-0.002	-0.014
		(0.017)	(0.019)	(0.021)
Panel C: High Propensity Youth in HH				
Treated in HH	0.005	0.021	0.012	-0.006
	(0.012)	(0.023)	(0.021)	(0.015)
Controls	Yes	Yes	Yes	Yes
Control group $\bar{Y}$	0.631	0.347	0.561	0.797
Observations	9.077	1.305	3.924	3.848
		Marginal - Low	Marginal - High	High - Low
Differences		-0.009	0.017	-0.027
		(0.031)	(0.025)	(0.028)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

Table A17: Spillovers by Combination of Voting Propensity Groups and Voted in 2021

	Voted			
	All	Low Propensity {Bottom 25%}	Marginal Voters {25-75%}	High Propensity {Top 25%}
	(1)	(2)	(3)	(4)
Panel A: Voted in 2021				
Treated	0.028***	0.062***	0.030***	0.009
	(0.007)	(0.019)	(0.012)	(0.010)
Controls	Yes	Yes	Yes	Yes
Untreated $\bar{Y}$	0.594	0.431	0.579	0.692
Observations	17,594	2,871	8,769	5,954
		Marginal - Low	Marginal - High	High - Low
Differences		-0.031	0.021	-0.052**
		(0.022)	(0.015)	(0.021)
Panel B: Did Not Vote in 2021				
Treated	0.006	0.009*	0.012**	-0.016
	(0.004)	(0.005)	(0.006)	(0.010)
Controls	Yes	Yes	Yes	Yes
Untreated $\bar{Y}$	0.115	0.057	0.118	0.210
Observations	27,678	8,789	14,177	4,712
		Marginal - Low	Marginal - High	High - Low
Differences		0.003	0.028**	-0.025**
		(0.008)	(0.012)	(0.011)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time. The analyses reported in the table were not pre-registered and are added ex post.

A.1.6 Number of Activated Voters

Table A18 shows number of activated voters by vote propensity groups for direct effect and spillover effect samples, and by age for the spillover sample. Where the definition of number of activated voters is $\beta_1 Treatment \times N_{treated\_individuals}$ for a given sample.

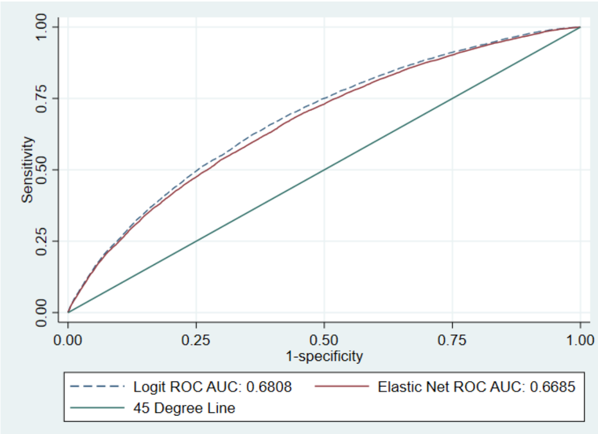
Table A18: Number of Activated Voters

	All	Low Propensity {Bottom 25%}	Marginal Voters {25-75%}	High Propensity {Top 25%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Youth Activated Voters	268.2	149.5	177.3	-59.2
Panel B: Spillover Effects				
Activated Voters	288.8	115.9	176.7	-33.4
Youth Activated Voters	53.6			
Over 29 Years Old Activated Voters	232.9			

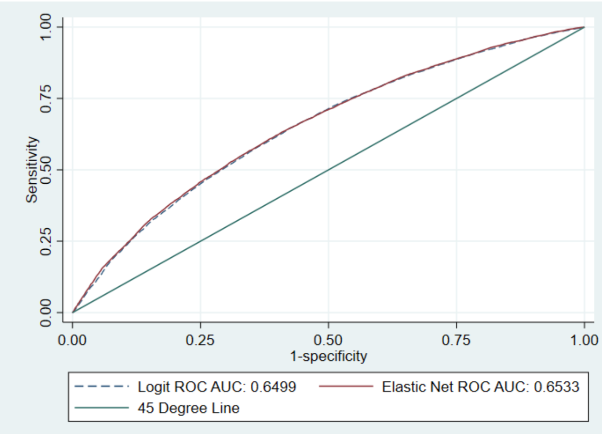
Notes: Activated voters defined as the treatment point estimate times the number of treatment group individuals in each group. Voting propensity groups and estimates from the logit prediction model. The analyses reported in the table were not pre-registered and are added ex post.

A.2 Supplementary Figures

A.2.1 Receiving Operating Characteristic (ROC) Curves



(a) ROC Curves for In-Sample Prediction



(b) ROC Curves for Out-of-Sample Prediction

Notes: For out-of- sample prediction sample is randomly split in order to estimate the model in the other half of the sample and compute the prediction error in the other half of the sample.

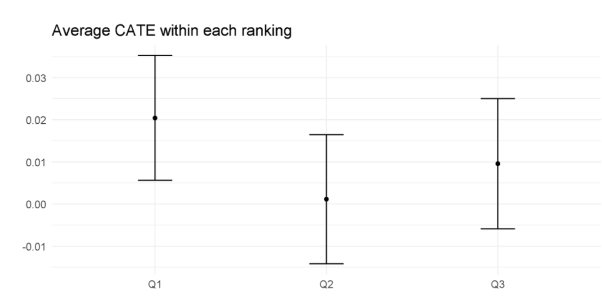
Figure A1: Receiving Operating Characteristic (ROC) Curves

A.2.2 Heterogeneous effects by honest causal forest

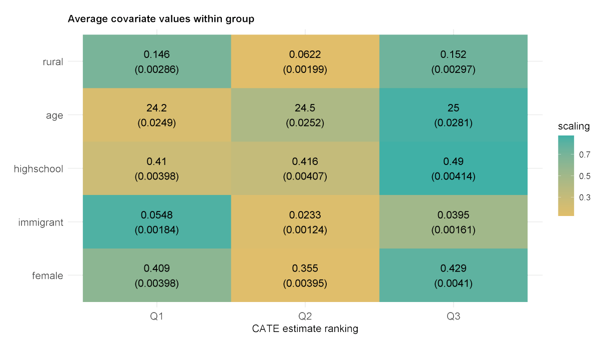
In this section, we employ a machine learning method to further assess the potential heterogeneity of treatment effects and their consequences for social inequalities in voting. The honest causal forest approach explores the heterogeneity of treatment effect using a multi-step procedure to avoid over-fitting the data. The honest causal forest method partitions sample according to splits by covariates into leafs and estimates conditional average treatment effects in each of these leafs. This splitting procedure is repeated many times to find which splittings lead to consistently larger differences in the treatment effects giving conditional treatment effect prediction for each observation. Observations are ranked by their conditional treatment effect prediction and quantiles are formed to compare covariate means across the predicted treatment effects. The honest causal forest algorithm separates the splitting and estimation of the conditional average treatment effect by using part of the sample for the former task and another part for the latter. The advantage of this method is that we do not need to assume at which dimensions the treatment effect heterogeneity takes place, which could be difficult to do based on theory ex-ante. The drawback is that a smaller sample is used for the actual estimation leading to noisy estimates. For this reason and given our limited sample size, we do not consider this analysis to be as revealing or important as the two approaches above.

Figure A2 (Panel A) shows that, using the honest causal forest algorithm, there is no evidence for treatment effect heterogeneity for direct treatment effect estimates. In fact, the first group, which is predicted to have the lowest conditional average treatment effect (CATE) in the training sample, has the highest point estimate in the evaluation sample and none of the estimates are statistically different from each other.

Figure A2 (Panel B) shows that the highest CATE ranking group (colors scaled as 0.5 being the mean) has the highest mean of educational attainment and the highest proportion of females. Individuals in this group are also on average slightly older compared to the individuals in the two other groups. However, observed demographic differences between the groups are not large, as expected as we do not observe differences in CATE estimates between the groups. We interpret these observations as further evidence that the intervention did not exacerbate inequality in turnout among the young voters. However, the lack of heterogeneous effects in this type of analysis is likely a result of low statistical power that results from dividing the sample by folding, and thus, does necessarily contradict with the previously documented, and more precisely estimated, results.



(a) Conditional Average Treatment Effect Estimates

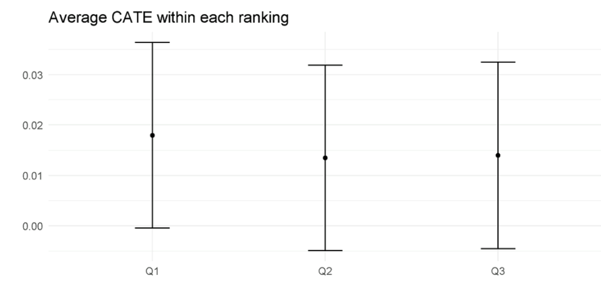


(b) Composition of CATE Ranking Groups

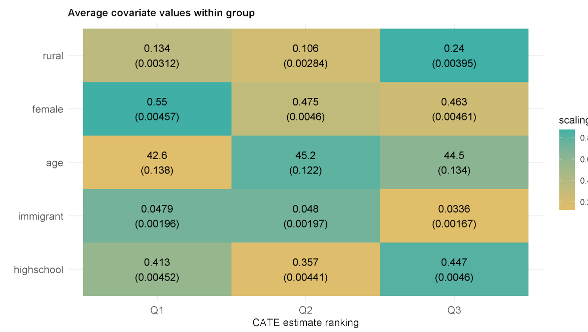
Notes: For Panel B colors scaled as 0.5 being the mean of the covariate in the CATE ranking group. The analyses reported in the figure were not pre-registered and are added ex post.

Figure A2: Honest Causal Random Forest Estimates - Direct Effects

Figure A3 (Panel A) shows that there are no statistically significant differences between the CATE ranking groups for spillover effect estimates. The highest CATE group has 24.0% individuals living in rural municipalities compared to 13.4% and 10.6% in the lowest and in the middle groups, respectively. The proportion of women in the lowest CATE group is 55.0%, whereas the proportion of women in the middle group is 47.5% and the proportion of women in the highest group is 46.3%. For other covariates, the differences between group means are smaller. Overall, using the honest causal forest algorithm, we find little evidence for large treatment effect heterogeneities in spillovers to untreated household members.



(a) Conditional Average Treatment Effect Estimates



(b) Composition of CATE Ranking Groups

Notes: For Panel B colors scaled as 0.5 being the mean of the covariate in the CATE ranking group. The analyses reported in the figure were not pre-registered and are added ex post

Figure A3: Honest Causal Forest Estimates - Spillover Effects

The **Aboa Centre for Economics (ACE)** is a joint initiative of the economics departments of the Turku School of Economics at the University of Turku and the School of Business and Economics at Åbo Akademi University. ACE was founded in 1998. The aim of the Centre is to coordinate research and education related to economics.

Contact information: Aboa Centre for Economics, Department of Economics, Rehtorinpellonkatu 3, FI-20500 Turku, Finland.

www.ace-economics.fi

ISSN 1796-3133