Political Behavior
https://doi.org/10.1007/s11109-024-09954-6

ORIGINAL PAPER

Who is Mobilized to Vote by Short Text Messages? Evidence from a Nationwide Field Experiment with Young Voters

Salomo Hirvonen[1] · Maarit Lassander[2] · Lauri Sääksvuori[3] · Janne Tukiainen[1]

Accepted: 21 June 2024

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. As a result, the GOTV intervention reduced existing social inequalities in voting within the target group of young voters. We remarkably find that over 100% of the direct treatment effect spilled over to untreated household members. These spillovers reduced inequality in political participation among the older voters who 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.

Keywords Field experiments · Get-out-the-vote · Inequality in voting · Spillover effects · Voter turnout

We thank the editor and the anonymous reviewers, Kasper M. Hansen, and the seminar participants at APSA, EPSA, FEAAM, the University of Helsinki, Laurea, IMEBESS and SABE/IAREP for helpful comments. The experiment was conducted in collaboration with the Ministry of Justice (Finland) and the Prime Minister's Office (Finland). We thank these organizations for funding and support. The authors have no relevant financial or non-financial interests to disclose. This RCT was registered in the American Economic Association's Registry for randomized controlled trials (RCT ID: AEARCTR-0008790). The Ethics Committee for Human Sciences at the University of Turku approved this study (48/2021).

✉ Janne Tukiainen
janne.tukiainen@utu.fi

1 Department of Economics, University of Turku, Rehtorinpellonkatu 3, 20014 Turku, Finland

2 Prime Minister's Office, Snellmaninkatu 1, 00023 Helsinki, Finland

3 Finnish Institute for Health and Welfare, Mannerheimintie 166, 00271 Helsinki, Finland

Published online: 10 July 2024

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 existing 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

[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.

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. 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 & Nickerson, 2009). However, to date, there is relatively little evidence on the compositional effects of text message-based mobilization strategies. 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 & Christensen, 2024), we find that our intervention mobilized mainly low-propensity voters. More uniquely, we find that the effects are the largest for the citizens belonging to low-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 & Christensen, 2024). We find a statistically significant, about 0.9% point, direct average treatment effect in the probability of voting. The effect size of 0.9% 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 & Gerber, 2019; Bergh & Christensen, 2024; Mann & 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 & Gerber, 2019). We find the largest effects for neutral message types.

The paper proceeds as follows. In Sect. "Theoretical Background", we discuss theoretical arguments regarding the potential effectiveness of the intervention. Section "Background and Context" describes the relevant electoral context. In Sect. "Experimental Design and Data", we describe our data, experimental design and the sample. Section "Data Analysis" presents our empirical methods. Section "Results" presents the results. Section "Discussion" discusses the results before Sect. "Conclusions" concludes. The paper is accompanied by an online supplement.

Theoretical Background

The most prominent theory regarding the average treatment effects of text message reminders on turnout is the Noticeable Reminder Theory (Dale & 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 & 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 an intention to vote and yet fail to do so.

The Receive-Accept-Sample theory (Zaller, 1992; Matland & 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 & 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 & 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 spillover effects 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.

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 & 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 5 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% points as women had a turnout of 40.7% and men had a turnout of 32.7%.[3]

[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 years) and older voters (over 29 years of age) in these elections was around 11% points. In the 2019 elections, the gender gap within the group of young voters

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

Experimental Design and Data

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 51,101 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) and (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

Footnote 3 (Continued)

was around 10% points. The gender gap in turnout remains similar after controlling for differences in educational attainment between men and women.

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

Table 1 Summary statistics: sample compared to population

	Analysis sample Full sample	Analysis sample Aged 19 to 29	Analysis municipalities Aged 19 to 29	Full population Aged 19 to 29
	(1)	(2)	(3)	(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	158,781	15,808	13,539	13,972
	(13,163)	(13,160)	(12,399)	(12,553)
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

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)

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 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.

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% of individuals into a control group and 60% of individuals into

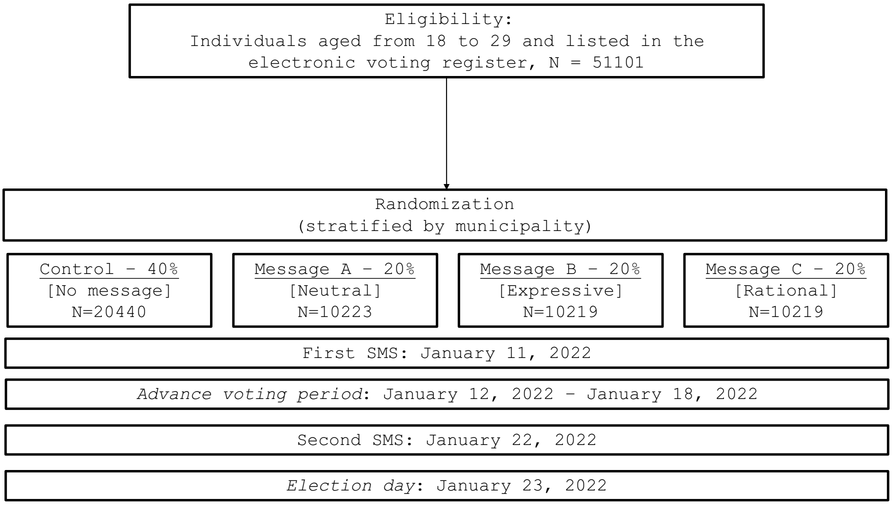

Fig. 1 Eligibility, randomization and treatment

three equally sized treatment groups (Fig. 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.

Message Contents

Since the popularization of the nudge theory (Thaler & 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 & Strauss, 2009) implies that the content of text messages should not affect turnout. However, there is still little 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 instrumental motivations to vote. The second type of messages was developed to appeal to the expressive motive of voting (Brennan & 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 & 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). At the same time, many citizens are averse to paternalism in other contexts (Lassen & Mahler, 2023). We conjecture that non-neutral message content may be considered paternalistic, which would imply 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 to 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 to 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 to 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]

Data Analysis

Using the RCT design and rich individual-level data on eligible voters and their household members, we conduct three different types of empirical analyses. The following section describes how these data analyses were carried out in our study. We note that we registered a pre-analysis plan (PAP) for our study by completing the registration fields in the American Economic Association's RCT Registry before implementing the study (Banerjee et al., 2020). Our PAP serves as a concise record of our initial intentions about the sample size, key outcomes and planned statistical analyses. However, our PAP did not provide a detailed description of the statistical analyses included in this article and did not specify in advance how the results are presented. Consequently, the statistical analyses presented in this paper do not follow a detailed PAP and should be primarily treated as empirical analyses that were conducted without a pre-registered statistical analysis plan.

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. 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

[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.

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]

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 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.

[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.

[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.

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, 25–75th, and top 25th percentiles.[11] This grouping is done to detect possible non-linear effects by voting propensity (Arceneaux & 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]

[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 A8) provides means of used covariates by voting propensity group for the main and spillover samples.

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 & 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) sub-samples 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 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 a researcher in terms of model specification. The results from the honest causal forest approach are presented only in the Online Appendix.

Results

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% 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%. Thus, the effect size of 0.9 p.p. equals around 3% increase compared to the

[13] We report in the Online Appendix (Fig. 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.

Table 3 Average treatment effect

	Voted				
	(1)	(2)	(3)	(4)	(5)
Treatment (pooled)	0.009*** (0.003)	0.009** (0.003)	0.009*** (0.003)	0.009** (0.003)	0.009*** (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	0.307	0.308
Observations	50.140	49.679	49.679	50.140	49.679

Statistical significance shown as ***p < 0.01, **p < 0.05, *p < 0.1, standard errors clustered at the municipal level in parentheses

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 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]

[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

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)

Statistical significance shown as ***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

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% 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 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. 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

Footnote 14 (Continued)

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 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

Statistical significance shown as ***p < 0.01, **p < 0.05, *p < 0.1, standard errors clustered at the municipal level in parentheses

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)	

Statistical significance shown as ***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 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

[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.

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 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.

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 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.

[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, 2022).

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

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)

Statistical significance shown as ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, 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

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 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 & 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.

[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 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)

Statistical significance shown as ***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

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

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)	

Statistical significance shown as ***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

methods. These insights, however, may hold limited practical implications for the development of concrete strategies to ameliorate gaps in political participation as many organizations implementing GOTV interventions 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). 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. It is noteworthy that 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 present 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.) 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, in the Online Appendix Table A19 we show results for sample splits by income and profession skill groups. In line with our main heterogeneous results, we observe that treatment effect is higher for individuals with low income and low skilled profession backgrounds.

Furthermore, we employ a more data-driven machine learning approach for the estimation of potential heterogeneous treatment effects. Online Appendix (Figs. A2 and 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.

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. 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

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

[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.

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 the freedom of a 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 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 & 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.

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 & Niemi, 1981; Linimon & 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 & 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.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11109-024-09954-6.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital).

Data availability Code and log-files for results stored at https://doi.org/10.7910/DVN/PGSVJP, Harvard Dataverse.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2022). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics, 138*(1), 1–35.

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

Banerjee, A., Duflo, E., Finkelstein, A., Katz, L. F., Olken, B. A., & Sautmann, A. (2020). In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics. Technical Report, National Bureau of Economic Research.

Bergh, J., & Christensen, D. A. (2024). 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, 34*(1), 79–95.

Bergh, J., Christensen, D. A., & Matland, R. E. (2020). Inviting immigrants in: Field experiments in voter mobilization among immigrants in Norway. *Electoral Studies, 66*, 102160.

Bergh, J., Christensen, D. A., & Matland, R. E. (2021). When is a reminder enough? Text message voter mobilization in a European context. *Political Behavior, 43*(3), 1091–1111.

Bhatti, Y., Dahlgaard, J. O., Hansen, J. H., & Hansen, K. M. (2017a). How voter mobilization from short text messages travels within households and families: Evidence from two nationwide field experiments. *Electoral Studies, 50*, 39–49.

Bhatti, Y., Dahlgaard, J. O., Hansen, J. H., & Hansen, K. M. (2017b). 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, 27*(3), 291–310.

Bhatti, Y., Dahlgaard, J. O., Hansen, J. H., & Hansen, K. M. (2018). Can governments use Get Out The Vote letters to solve Europe's turnout crisis? Evidence from a field experiment. *West European Politics, 41*(1), 240–260.

Blais, A., Young, R., & Lapp, M. (2000). The calculus of voting: An empirical test. *European Journal of Political Research, 37*(2), 181–201.

Brennan, G., & Brooks, M. (2013). Expressive voting. In *The Elgar companion to public choice* (2nd ed.). Edward Elgar Publishing.

Cheng-Matsuno, V., Foos, F., John, P., & Unan, A. (2023). Do text messages increase voter registration? Evidence from RCTs with a local authority and an advocacy organisation in the UK. *Electoral Studies, 81*, 102572.

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

Dahlgaard, J. O., Bhatti, Y., Hansen, J. H., & Hansen, K. M. (2022). Living together, voting together: Voters moving in together before an election have higher turnout. *British Journal of Political Science, 52*(2), 631–648.

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

Downs, A. (1957). *An economic theory of democracy*. Harper.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics, 4*, 3895–3962.

Enos, R. D., Fowler, A., & Vavreck, L. (2014). Increasing inequality: The effect of GOTV mobilization on the composition of the electorate. *The Journal of Politics, 76*(1), 273–288.

Foos, F., & de Rooij, E. A. (2017). All in the family: Partisan disagreement and electoral mobilization in intimate networks-a spillover experiment. *American Journal of Political Science, 61*(2), 289–304.

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

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

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review, 102*(1), 33–48.

Green, D. P., & Gerber, A. S. (2019). *Get out the vote: How to increase voter turnout*. Brookings Institution Press.

Green, D. P., McGrath, M. C., & Aronow, P. M. (2013). Field experiments and the study of voter turnout. *Journal of Elections, Public Opinion and Parties, 23*(1), 27–48.

Harjunen, O., Saarimaa, T., & Tukiainen, J. (2023). Love thy (elected) neighbor? Residential segregation, political representation, and local public goods. *The Journal of Politics, 85*, 860–875.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity. *Monographs on Statistics and Applied Probability, 143*, 143.

Hirvonen, S., Schafer, J., & Tukiainen, J. (2023). *Policy feedback and civic engagement: Evidence from the Finnish basic income experiment. Discussion paper 155*. Aboa Centre for Economics.

Jääskeläinen, A. (2020). *The Finnish election system: Overview*. Oikeusministerio.

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

Kefford, G., Dommett, K., Baldwin-Philippi, J., Bannerman, S., Dobber, T., Kruschinski, S., Kruikemeier, S., & Rzepecki, E. (2023). Data-driven campaigning and democratic disruption: Evidence from six advanced democracies. *Party Politics, 29*(3), 448–462.

Klofstad, C. A. (2007). Talk leads to recruitment: How discussions about politics and current events increase civic participation. *Political Research Quarterly, 60*(2), 180–191.

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

Lim, S. S., Updike, R. L., Kaldjian, A. S., Barber, R. M., Cowling, K., York, H., Friedman, J., Xu, R., Whisnant, J. L., Taylor, H. J., et al. (2018). Measuring human capital: A systematic analysis of 195 countries and territories, 1990–2016. *The Lancet, 392*(10154), 1217–1234.

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

Lyytikäinen, T., & Tukiainen, Janne. (2019). Are voters rational? *European Journal of Political Economy, 59*, 230–242.

Malhotra, N., Michelson, M. R., Rogers, T., & Valenzuela, A. A. (2011). Text messages as mobilization tools: The conditional effect of habitual voting and election salience. *American Politics Research, 39*(4), 664–681.

Mann, C. B., & Haenschen, Katherine. (2024). A meta-analysis of voter mobilization tactics by electoral salience. *Electoral Studies, 87*, 102729.

Matland, R. E., & Murray, G. R. (2012). An experimental test of mobilization effects in a Latino community. *Political Research Quarterly, 65*(1), 192–205.

Mo, C. H., Holbein, J. B., & Elder, Elizabeth Mitchell. (2022). Civilian national service programs can powerfully increase youth voter turnout. *Proceedings of the National Academy of Sciences, 119*(29), e2122996119.

Naess, O.-A. E. (2022). Increasing turnout with a text message: Evidence from a large campaign from the government. *Journal of Elections, Public Opinion and Parties, 34*, 212–230.

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

Sinclair, B., McConnell, M., & Green, D. P. (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science, 56*(4), 1055–1069.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

von Schoultz, Å., & Strandberg, K. (2024). *Political behaviour in contemporary Finland: Studies of voting and campaigning in a candidate-oriented political system*. Routledge.

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

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301–320.

Zuckerman, A. S. (2005). *The social logic of politics: Personal networks as contexts for political behavior*. Temple University Press.