Journal of Experimental Political Science (2023), **10,** 188–200
doi:10.1017/XPS.2022.16

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Large-Scale Evidence for the Effectiveness of Partisan GOTV Robo Calls

Daniel T. Kling[1]* and Thomas Stratmann[2]

[1]Department of Business, Belmont Abbey College, Belmont, NC, USA and [2]Department of Economics, George Mason University, Fairfax, VA, USA
*Correspondence: Email: dan.kling@gmail.com

Abstract

We document the effectiveness of automated (robo) calls for increasing voter participation in contrast to most published research which finds little or no effect from automated calls. We establish this finding in a large field experiment which mimics campaign behavior with a targeted, partisan get-out-the-vote campaign. Our findings show that across all treatments, automated calls led to three additional votes for every thousand subjects called during the 2014 midterm general election. Additionally, our experimental design allows for testing how the number of calls in a treatment, that is dosage, affects voter turnout. Here, results show that three extra calls increase the treatment effect to seven additional votes per thousand subjects called, but that too many additional calls decrease that effect to statistical insignificance in a six-call treatment.

Keywords: elections; voter turnout; robo calls; field experiment; GOTV

Introduction

We conducted a large-scale experiment to assess the effectiveness of partisan robo calls.[1] We find a positive, significant treatment effect of increasing voter turnout. Our experiment has several key features that stand out in the get-out-the-vote (GOTV) literature. First, our experiment varies the intensity of the treatment dosage by using a different number of calls (one, three, or six) for different treatment groups. We found the largest treatment effects for subjects scheduled to receive three calls. Second, a Republican-leaning media firm administered the experiment and targeted likely Republican voters. In contrast, many similar published experiments are nonpartisan, Democratic, or otherwise left-leaning. Third, the subject target of our experiment was potentially marginal voters. Specifically, people who had voted in two of the previous four elections. These

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

[1]The experiment itself was conducted by a political consulting firm while the authors of this paper performed the analysis of the data. In the description of the experiment and its execution, we intend for "we" and "our" to include the members of that firm.

© The Author(s), 2022. Published by Cambridge University Press on behalf of The Experimental Research Section of the American Political Science Association.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

three key features may explain why this experiment shows that robo calls have a positive, significant treatment effect on voter participation, while many other robo call experiments do not. In this section, we will discuss important results from the robo call GOTV literature and highlight several important features of the experiment and how they relate to previous GOTV studies.

Previous research shows that the impact of robo calls on voter turnout is small and not statistically significant. This is summarized in a meta-study by Green, McGrath, and Aronow (2013). They report a statistically insignificant increase in voter turnout of 0.156 percentage points, attributable to prerecorded phone messages. Another meta-analysis reports an "average intent-to-treat effect of 0.113 percentage point, with a 95% confidence interval ranging from −0.336 to 0.563" (Green and Gerber 2015, p. 196), also implying a statistically insignificant effect of robo calls on voter participation.[2]

One study that did find a significant positive treatment effect of robo calls is Gerber et al. (2010), which used a placebo design to increase the precision of the estimate of its treatment effect.[3] Studying the 2008 primary election, Gerber et al. (2010) find an approximately two percentage point increase in voter participation for subjects who were successfully contacted, relative to subjects in a placebo group who were successfully contacted. The authors attribute the large size of their reported effect to call content, or "social pressure" messaging, which reminded subjects that they had voted in the past two general elections but not in the past primary, and informed subjects that their voting records are publicly available. However, when attempting to replicate their results for the November 2008 general election in a follow-up study, the authors find "a weakly positive effect in the November 2008 general election, in keeping with the usual pattern of weaker turnout effects in high-salience elections" (Green and Gerber 2019, 89). Ali and Lin (2013) develop a model of voter turnout based on similar social dynamics. DellaVigna et al. (2014) present evidence from a field experiment that "individuals vote because they expect to be asked" about it and find lying costly. That research suggests that understanding the dynamics of social pressure messaging amidst an evolving campaign landscape is a fruitful research agenda.

Our experiment does not use social pressure messaging but does find a statistically significant positive effect for robo calls placed before the November 2014 general election. Subjects in the combined treatment group were 0.318 percentage points more likely to vote than subjects in the control group – or a rate of 1 mobilized voter for every 314 subjects called. The treatment with three calls was the most effective. Subjects in the three-call treatment were 0.695 percentage points more likely to vote than subjects in the control group – or a rate of 1 mobilized voter for every 144 subjects called.

We posit that robo calls are most effective when they are received by voters who are on the margin as to whether to participate or not to participate in voting. That is,

[2]A more recent meta-analysis (Green and Gerber 2019, p. 222) finds an average treatment effect of 0.234 percentage points (confidence interval from 0.039 to 0.430 percentage points). However, that newer meta-analysis does not provide the best of comparison because it includes the results from this experiment as described in a related article based on the experiment described in this paper.

[3]See also Nickerson (2008) for an early example a GOTV placebo design.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

a GOTV treatment will be less effective if it is primarily targeted to individuals who are habitual voters or individuals who never vote. Most of the individuals with a high propensity to vote will vote regardless of the treatment. Similarly, individuals with a very low propensity to vote are unlikely to be mobilized by a robo call. However, an individual's propensity to vote is context-dependent and can change based on the salience of a given election to that potential voter (Arceneaux and Nickerson 2009). Moreover, that an individual's propensity to vote is context dependent is consistent with an array of rational choice models of voting (Ali and Lin 2013; Coate and Conlin 2004; Degan and Merlo 2011; Dhillon and Peralta 2002; Feddersen and Sandroni 2006). Thus, along several dimensions, such as the choice of states without multiple high-profile statewide races and the subject selection criteria, the experiment was designed to include a higher proportion of subjects who were likely to be marginal voters.

We built a sample of potentially marginal voters by focusing on registered voters who had voted exactly twice, during the 2010 and 2012 primary and general elections. Thus, these voters missed voting in two of the four elections. Other GOTV studies use alternative criteria for subject selection. For example, Ramirez (2005, p. 70) selects subjects from a list of "registered Latino voters in low-propensity precincts" Other studies select subjects that are expected to be receptive to the treatment message, which is in a similar spirit to our experimental design. For example, Shaw et al. (2012, p. 236), whose calls include an endorsement by Texas Governor Rick Perry, direct their calls to subjects that are "both likely primary voters and strong Perry supporters." Gerber et al. (2010) select only subjects who voted in the past two general elections but did not vote in the most recent primary.

As noted above, our experiment is targeted at Republican voters, while many other robo call and other GOTV experiments target Democratic or nonpartisan voters. Examples of such experiments include, e.g., Nickerson and Rogers (2010), Barton, Castillo, and Petrie (2016), and Rogers, Green, Ternovski, and Young (2017) which are all targeted to Democrats (with messages of varying degrees of partisanship) while Ramirez (2005), Nickerson (2008), Gerber, Green, Kaplan, and Kern (2010), and Panagopoulos (2011) had nonpartisan targeting. Among studies of robo calls, Shaw, Green, Gimpel, and Gerber (2012) is an exception; it targeted voters in a Republican primary.

It is possible that differences between partisan groups or differences in messages could cause particular GOTV methods, such as robo calls, to be more effective for one voter group than another. This effect could possibly be mediated by voters' age or other voter characteristics. We do not have prior expectations that robo calls are more (or less) effective when targeting Republican voters as opposed to Democratic or nonpartisan voters, especially after controlling for individual characteristics. However, we believe that it is important to close the empirical gap in studies of the effectiveness of robo calls with experiments that have Republicans in general elections as their main subject pool. This approach will give insights into whether findings from experiments targeting Democratic and nonpartisan voters generalize to Republican voters.

One of the important features of this experiment is the use of treatment groups that vary in the number of GOTV robo calls received. With a few exceptions, published GOTV experiments use a single robo call, while it is common for political

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

candidates and advocacy groups to deploy multiple iterations of multiple types of GOTV communication. Ramirez (2005) included a treatment with two robo calls among a wide assortment of GOTV methods. Recently, Zelizer (2020) studied the effect of repeating robo calls, dividing approximately 40,000 registered voters into eight treatments that received zero to seven calls in the week leading up to a primary election. Zelizer found that a treatment with five calls was the most effective, but the differences between treatments were not statistically significant. We are not aware of any other studies that directly measure the effect of robo call dosage.

Similarly, Green and Zelizer (2017) conducted an experiment using multiple treatments with GOTV mail and found that efficacy declined after five mailers. Other experimental studies have raised the issue of GOTV over-saturation (Gerber and Green 2000) or analyzed GOTV synergies via multiple treatments (Green and Gerber 2019, p. 184), but did not find an effect of such multiple treatments. In a manual for political operatives, Lofy (2005, p. 153) recommends two to nine contacts before an election to increase the likelihood that habitual voters cast a ballot on the day of the election.

Procedures

The field experiment occurred in the days leading up to the November 2014 general election. The experiment was funded and conducted by a political consulting firm seeking to evaluate the effectiveness of their GOTV services.[4] The firm identified registered voters in six states as subjects. The potential subject pool consisted of likely Republican voters who had registered to vote in Georgia, Nebraska, New Mexico, Ohio, Pennsylvania, or Virginia prior to January 1, 2010.[5] The pool of potential subjects was narrowed to people who had voted in exactly two of the 2010 and 2012 primary and general elections. People who voted early or who cast an absentee ballot in the 2010 or 2012 general election were also excluded.

From this group, 42,000 subjects with unique landline phone numbers were randomly selected from each state. Many of these landline numbers are associated with households that have more than one voter. Because we cannot identify who answers a robo call, we included every registered voter in each selected household in the analysis even though these household members are not expected to all match the original selection criteria in terms of voting history or even party affiliation. The 42,000 households from each state yielded subject pools ranging from 86,714 to 95,557 registered voters. In total, the study includes 539,567 subjects.

[4]This experiment was not pre-registered. The authors received approval to perform this research from their institution's Institutional Review Board. Initially, our involvement with the experiment was limited to providing information about design features for clean causal inference, identifying the voters likely to be marginal, and the sample size necessary to detect an expected treatment effect. We were subsequently asked to assist with analyzing and interpreting the data. We received permission to use the data for academic research and publication a year after the experiment.

[5]In Nebraska, New Mexico, and Pennsylvania, voters register with a party affiliation when they register to vote. Georgia, Ohio, and Virginia do not have party registration. For these three states, the likely party affiliation and eligibility for the subject pool were determined by a third-party data company. We expect that the results from this experiment would generalize to voters from either of the two parties.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

Within each state, households were randomly assigned to one of four groups: a control group and three treatment groups. Subjects in the three treatment groups, T1, T3, and T6 received one, three, or six treatment calls over 1, 3, or 6 days, respectively. The robo calls to subjects in T6 started 6 days before the election, the calls to subjects in T3 started 3 days before the election, and all subjects in the three treatment groups received a call on the morning of Election Day. On average, each household has 2.16 registered voters. Detailed information, including summary statistics of subject characteristics in each treatment group, is in Appendix A. The table of summary statistics shows that there are some imbalances across treatments on some control variables, particularly 2010 and 2012 voting rates. To address this issue, we include these covariates in our regression models. Further, we cluster the standard errors by household in all household-level specifications.

Treatment call messages were slightly different on each call date. Each message included a reminder about the date of the election and a short partisan message encouraging the subject to vote. Most of the messages invoke negative partisanship rather than making specific promises or describing specific plans.[6] The duration of the messages was between 35 and 45 seconds. Appendix B lists the script for each message. Table 1 shows the schedule of calls and schedule of scripts, which were identical for every state.

We define "live answerers" as subjects residing in households where at least one call resulted in a live answer, meaning someone in that household answered the phone for any length of time. We classify all other subjects as "treatment non-answerers," regardless of whether the call went to an answering machine, operator, no answer, busy, fax machine, or was otherwise uncompleted.[7] However, calls that went to answering machines or had a live answer were considered successful treatment for the purposes of the local average treatment effect (LATE) analysis described in Appendix E, because subjects may have listened to and been influenced by the message left on their answering machine. Our outcome variable – whether a subject voted in the election – is based on verified voting records for the November 2014 election.

Results

Table 2 shows that the percent of subjects with at least one live answer for T1, T3, and T6 were 39, 60, and 67%, respectively.[8] The remaining subjects in each group did not answer any of the treatment calls. Table 3 shows the mean number of calls

[6]Negative messages are often more effective for motivating voters (Barton, Castillo, and Petrie 2016).

[7]If a call outcome was operator, no answer, busy, fax machine, or otherwise uncompleted on the first attempt, then the phone number was called again 30 minutes later. If the second attempt also was not successful, a final attempt was made after another 30 minutes had passed. Thus, each single treatment "call" could include up to three attempts. Once a call resulted in a live answer or answering machine on the day for which the call was scheduled according to the treatment protocol, on that day, the message was played, and no further calls were placed. In all cases, the outcome of the first successful or the final unsuccessful call was recorded as the outcome of that call.

[8]We report detailed results of the treatment call outcomes in Appendix C. That appendix includes call outcomes grouped by state, call number, and treatment group.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

Table 1.
Treatment Call Schedule

Group	Oct 30	Oct 31	Nov 1	Nov 1	Nov 3	Nov 4
Control	–	–	–	–	–	–
T1	–	–	–	–	–	Script 6
T3	–	–	–	Script 4	Script 5	Script 6
T6	Script 1	Script 2	Script 3	Script 4	Script 5	Script 6

Note: Households in the treatment groups were called in 2014, on the date listed in the first row of Table 3. The general election was on November 4, 2014. Appendix B includes the text for each of the six scripts.

Table 2.
Treatment Call Outcomes by Treatment Group

Call Outcomes	T1	T3	T6	Total
At least one live answer (Live answerer)	39%	60%	67%	55%
	(53,355)	(80,297)	(89,632)	(223,284)
No live answer	61%	40%	33%	45%
	(82,565)	(54,529)	(43,858)	(180,952)
At least one AM	34%	49%	56%	46%
	(45,995)	(65,689)	(74,347)	(186,031)
At least one AM, no live answers	34%	16%	9%	20%
	(45,995)	(21,078)	(12,556)	(79,629)
At least one live answer or AM (successful treatment)	73%	75%	77%	75%
	(99,350)	(101,375)	(102,188)	(302,913)
No live answer or AM	27%	25%	23%	25%
	(36,570)	(33,451)	(31,302)	(101,323)
Total	100%	100%	100%	100%
	(135,920)	(134,826)	(133,490)	(404,236)

Note: The table shows the percent of subjects in each treatment group that either answered at least one call live, had no live answer, had at least one answering machine answer, etc. We refer to these "responses" as "call outcomes." Treatment call outcome categories, which are not all mutually exclusive, are listed in the rows of this table. Percents are relative to the total number of subjects in each treatment group. "AM" denotes answering machine. The raw number of subjects in each cell is given in parentheses.

with live answers for subjects in each treatment group.[9] This pattern of more live answers in treatments with more calls is consistent across states.

[9]These tables show that the percent of live answerers for T3 was 21 percentage points higher than for T1, and live answerers in T3 answered one more call on average. However, the percent of live answerers for T6 was only 7 percentage points higher than for T3, while live answerers in T6 answered almost two more calls than live answerers in T3.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

Table 3.
Mean Number of Live Answers by Treatment Group

Mean number of live answers	T1	T3	T6	Total
Total treatment group	0.4	1.2	2.5	1.4
	(0.5)	(1.2)	(2.3)	(1.7)
Number of subjects	135,920	134,826	133,490	404,236
Live answerers only	1.0	2.1	3.8	2.5
	(0.0)	(0.8)	(1.7)	(1.6)
Number of subjects	53,355	80,297	89,632	223,284

Note: The top panel of this table shows the mean number of live answers for all subjects in each treatment group. The bottom panel shows the mean number of live answers only among subjects with at least one live answer. Standard errors for each mean are listed in parentheses.

Table 4.
Voting Rates by Treatment Group

	Voting rate (%)	Number of subjects
Control	48.86	135,331
T1	49.08	135,920
T3	49.51	134,826
T6	48.94	133,490
All treatments	49.18	404,236

Table 4 reports the voting rates for each of the treatment groups. These results show that treatment was associated with modestly higher voting rates, particularly the three call treatments, with voting rates among all treated subjects and subjects in T3 approximately 0.3 and 0.6 percentage points higher, respectively.

Table 5 uses multivariate regressions to report the intent-to-treat estimates from comparing voting rates of the three treatment groups to those of the control group while controlling for subject-specific voting history, demographics, and socioeconomic characteristics.[10] Specifically, our control variables include whether or not a subject voted in the 2010 and/or 2012 general elections and subject age, gender, educational attainment, estimated income, state of residence, and the number of registered voters in their household.[11] The coefficients on each of the treatment

[10]The corresponding table without subject controls is available in Appendix E. These estimates are similar to the estimates reported in Table 5.

[11]We truncate the values for age and the number of subjects in a household to reduce the probability that outliers will lead to spurious results. The low and high age categories are 20 or younger and 90 or older, respectively, while the maximum variable for the number of subjects in a household is coded as 6 or higher. Data for control variables were not available for a small number of subjects. For these subjects, we imputed values based on the means of the rest of the sample. When a value was missing for an indicator variable, we added an indicator for "data unavailable" and coded the missing indicator as 99.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

Table 5.
Intent-to-Treat Effects for All Subjects and Single-Voter Households (SVH)

	(1)	(2)	(3)	(4)
Variables	voted	voted	voted	voted
Population	All subjects	All subjects	SVH	SVH
T1		0.00118		0.00472
		(0.00192)		(0.00513)
T3		0.00695***		0.00976*
		(0.00192)		(0.00514)
T6		0.00140		−0.00244
		(0.00194)		(0.00507)
All Treatments	0.00318**		0.00388	
	(0.00157)		(0.00416)	
Subject controls	Yes	Yes	Yes	Yes
Observations	539,567	539,567	54,849	54,849
R-squared	0.291	0.291	0.243	0.243
F-test	9026	8480	846.1	793.4
Prob > F	<0.001	<0.001	<0.001	<0.001

Note: The table shows OLS estimates for the difference in voting rates between subjects in the treatment group and subjects in the control group. The subset of subjects used to estimate each specification is shown with the "Population" label. Subject controls include whether or not a subject voted in the 2010 or 2012 general elections or both and subject age, gender, education, income, number of subjects in the household, and state of residence. Standard errors are clustered by household and displayed in parentheses.

***$p < 0.01$,
**$p < 0.05$,
*$p < 0.1$.

variables show the difference in voting rates, measured in percentage points, between the corresponding treatment group and the control group.[12] Standard errors are clustered by household and listed below each regression coefficient.

In Table 5, Column 1, the coefficient on the *All Treatments* variable is 0.00318, which indicates that voter participation for subjects in the pooled three treatment group is 0.318 percentage points higher than for subjects in the control group. This point estimate is statistically significant at the 5% level. The magnitude of this estimate indicates a yield of three additional voters for every 1,000 subjects who were called, or 314 subjects called per additional voter mobilized.[13] This effect is larger than the average robo call effect reported in Green and Gerber (2015, p. 196). This

[12]Although we use OLS to estimate treatment effects, our results are robust to nonlinear estimation techniques such as probit and logit regression models.

[13]Given that each landline phone number was associated, on average, with approximately 2.5 household members, this estimate implies that robo call treatments for 400 households, with an average of 3.5 calls per treatment, lead to three additional voters.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

result implies a cost of under $9 to induce a subject from not participating in the election to casting a ballot.[14]

Table 5, Column 2 reports estimation results by each treatment group. The findings indicate that the most effective treatment is T3, the treatment in which subjects receive an automated call on each of 3 days. T3's treatment effect is 0.695 percentage points which is statistically significant at the one percent level.[15] This estimate implies the turnout of one additional voter for every 144 subjects assigned to T3.[16] The treatment effect for groups T1 and T6 are positive, but smaller and statistically insignificant. The implied cost to induce an additional vote for only subjects in T3 is under $4.

Our findings show that calling on 3 days was more effective than calling on 1 day or over 6 days.[17] The Wald statistic for testing for the equality of the coefficients for T1 and T3 is 9.04 and is significant at the one percent level. The Wald statistic testing the equality of the coefficients for T3 and T6 is 8.21 and is statistically significant at the one percent level. Thus, for both tests, we reject the hypothesis of equal dosage effects. There is no statistically significant difference between the coefficients for T1 and T6.

In households with multiple registered voters, we do not know which subject or subjects answered the treatment call or calls. Because the treatment effect may be diluted in multi-voter household, we also analyze subjects in single-voter households, which accounted for approximately one tenth of all subjects.[18] In Table 5, Columns 3 and 4 provide those estimation results. Our data show a larger treatment effect for subjects in single-voter households. In Column 3, the average treatment effect for single-voter households is 0.388 percentage points, but it is not statistically significant at the ten percent level.

Table 5, Column 4 separates the results by treatment group and shows the estimated effect among single-voter households. The treatment effects for T1 and T3 are both positive, although only T3 is statistically significant at the ten percent level. The estimate for T6 is negative, but not statistically significant.

As noted in the Procedures section above, the treatment messages were different from day to day. Within T3 and T6, subjects differed in terms of which days had live answers, and thus which messages were heard. However, we were not able to identify any meaningful differences in turnout that can be attributed to differences

[14]We estimate the total cost of generating one additional voter through automated calls by multiplying the sum of the number of calls that resulted in a live answer or answering machine by $250/10,000 calls. We estimate the number of additional voters by multiplying the treatment coefficient, 0.00318, by the number of subjects in the treatment group, 404,236. We estimate the cost per additional voter by dividing the total cost by the number of additional voters.

[15]Following the methodology described in List et al. (2019), we evaluate the effect of correcting for multiple hypothesis testing. With this correction, the point estimate for T3 remains statistically significant at the one percent level. See Appendix F for details.

[16]Based on the average of 2.5 household members per landline, this corresponds to approximately 62 treated households, or 186 robo calls placed.

[17]This finding corroborates a conventional GOTV strategy of making "two or three contacts in the final weekend" (Lofy 2005, 153).

[18]In an experiment to measure how making specific plans with subjects can increase the effectiveness of GOTV contacts, Nickerson and Rogers (2010) find much larger treatment effects for single-voter households.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

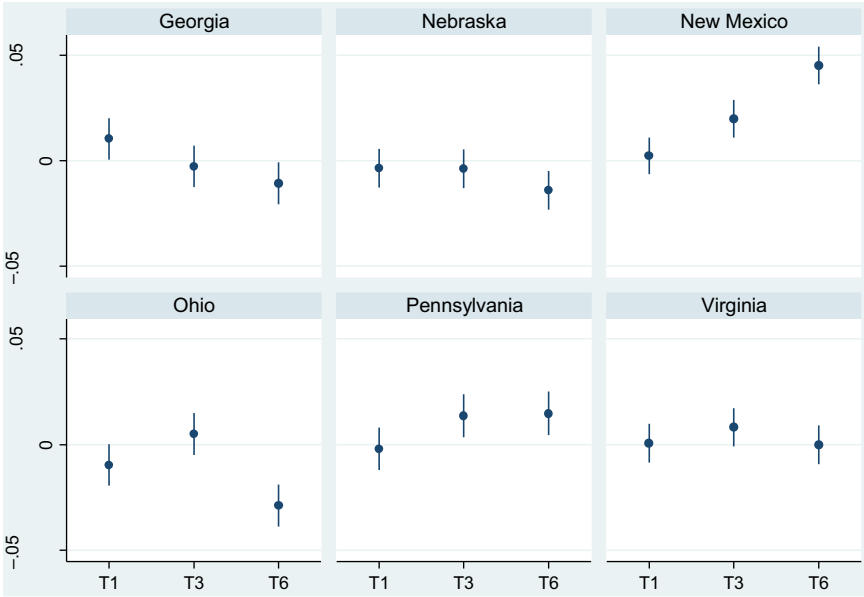

Figure 1.

Intent-to-treat effects by state and treatment group.

The chart shows the state-by-state OLS estimates for the difference in voting rates between subjects in each treatment group and subjects in the control group in that state. Each circle represents the estimated treatment effect, and the vertical lines indicate the 95% confidence interval for each estimate. The regression used to calculate these treatment effects used subject controls, including whether or not a subject voted in the 2010 or 2012 general elections or both and subject age, gender, education, income, and number of subjects in the household. Standard errors are clustered by household.

between hearing specific message scripts.[19] Similarly, although previous researchers find that GOTV calls are less effective the earlier these calls are placed prior to an election (Green and Gerber 2019, p. 186; Murray and Matland 2014; Nickerson 2007; Panagopoulos 2011), we did not observe a timing effect within the groups of voters who received three or six calls. This suggests that within the time frame of our study, the effectiveness of the treatment did not measurably decay.

We found that the pattern of treatment effectiveness varied across the states in our sample. We have included a chart in Figure 1 to show each group's treatment effect for the six states in the experiment. While there are differences in the balance of subject characteristics across states, we do not observe a correlation between differences in some of the balance tests and the estimated treatment effects across the states. The state-by-state subject characteristic summary statistics are included in Appendix A as are the regression tables for the results graphed in Figure 1.

In Appendix D, we discuss an alternative method of estimating the causal treatment effect using data from a call placed after the election. In Appendix E, we report

[19]For example, by restricting the sample to subjects with exactly one live answer we might be able to detect differences across messages or days, but we did not find anything conclusive to suggest such differences.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

the conventional method of estimating the treatment on treated effect by calculating local average treatment effects. In Appendix F, we test whether our results are significant after correcting for multiple hypothesis testing.

Conclusion

This experiment employed multiple-call dosage and mimicked the behavior of real political GOTV campaigns. We show that targeted automated calls have a positive effect on voter turnout even without using a social pressure message. We find that, across all treatments, the intent-to-treat effect is larger than most previous measures of robo call effectiveness. Our results suggest that it is not irrational for campaigns to deploy robo calls as a cheap additional tool to increase voter turnout. We find that the treatment increases voter participation by between three to six additional voters for every 1,000 subjects. This corresponds to a cost of $4 to $9 to induce a subject to vote in elections.

We also find that dosage matters. The intent-to-treat effect for dosage T3 is two to three times larger than for dosage T1, while dosage T6 typically had the smallest effect. These results suggest that additional calls can increase effectiveness, but that too many calls may be counterproductive. Our data are not rich enough to determine why the T6 treatment had the smallest treatment effect. One explanation for this finding might be that receiving so many calls annoyed the subjects and that they started ignoring the calls that were closer to the election, even if they were picking up the phone. While T6 reached a higher percentage of subjects than T3 – 67 and 60% – respectively, the increase from three calls to six calls nearly doubled the total number of live answers for each live answerer – from 2.1 in T3 to 3.8 in T6. Annoyance with this stark increase in live answers received may explain the decreased effectiveness of T6. This topic may warrant further experiments, using placebos, to determine how receiving marginal additional calls affects voting behavior.

Overall, the pattern of dosage results observed in our experiment is consistent with the broader GOTV literature. The 0.12 percentage point effect of T1 is consistent with meta-analyses (Green and Gerber 2015, p. 196; Green, McGrath, and Aronow 2013),[20] while the larger 0.65 percentage point treatment effect of T3 suggests, along with more recent work (such as Zelizer 2020), provides evidence that robo call treatments are more effective with some, but not too much, repetition.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/XPS.2022.16

Data availability. The data, code, and any additional materials required to replicate all analyses in this article are available at the Journal of Experimental Political Science Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/DMJ7EA.

Conflicts of interest. The authors are not aware of any conflicts of interest related to this paper.

[20]As noted above, the newer meta-analysis with a 0.234 percentage point treatment effect in Green and Gerber (2019, p. 222) includes experiments with multiple calls: Zelizer (2020) and the experiment described here.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

Ethics statement. The research presented here was approved by the George Mason University IRB under Project Title: [854034-1] Robo call GOTV data analysis.

To the best of our knowledge, this research adheres to APSA's Principles and Guidance for Human Subjects Research.

Details of the experimental procedure are described in the manuscript above. Further description pertinent to the APSA standards is in online Appendix G.

References

Ali, S. Nageeb and Charles Lin. 2013. Why People Vote: Ethical Motives and Social Incentives. *American Economic Journal: Microeconomics 5*(2): 73–98.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press.

Arceneaux, Kevin and David W. Nickerson. 2009. Who Is Mobilized to Vote? A Re-Analysis of 11 Field Experiments. *American Journal of Political Science 53*(1): 1–16.

Barton, Jared, Macro Castillo, and Ragan Petrie. 2016. Negative Campaigning, Fundraising, and Voter Turnout: A Field Experiment. *Journal of Economic Behavior & Organization*, *121*, 99–113.

Coate, Stephen and Michael Conlin. 2004. A Group Rule–Utilitarian Approach to Voter Turnout: Theory and Evidence. *The American Economic Review 94*(5): 1476–504.

Degan, Arianna and Antonio Merlo. 2011. A Structural Model of Turnout and Voting in Multiple Elections. *Journal of the European Economic Association 9*(2): 209–45.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2017. Voting to Tell Others. *The Review of Economic Studies 84*(1): 143–81.

Dhillon, Amrita and Susana Peralta. 2002. Economic Theories of Voter Turnout. *The Economic Journal 112*(480): F332–52.

Feddersen, Timothy J. and Alvaro Sandroni. 2006. A Theory of Participation in Elections. *The American Economic Review 96*(4): 1271–82.

Gerber, Alan S., and Donald P. Green. 2000. The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment. *American Political Science Review 94*(03): 653–63.

Gerber, Alan S., Donald P. Green, Edward H. Kaplan, and Holger L. Kern. 2010. Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments. *Political Analysis 18*(3): 297–315.

Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling, Connor Raso, and Shang E. Ha. 2011. Personality Traits and Participation in Political Processes. *The Journal of Politics 73*(3): 692–706.

Green, Donald P., and Alan S. Gerber. 2015. *Get Out the Vote: How to Increase Voter Turnout* (3rd ed.). Washington, DC: Brookings Institution Press.

Green, Donald P. and Alan S. Gerber. 2019. *Get Out the Vote: How to Increase Voter Turnout* (4th ed.). Washington, DC: Brookings Institution Press.

Green, Donald P., Mary C. McGrath, and Peter M. Aronow. 2013. Field Experiments and the Study of Voter Turnout. *Journal of Elections, Public Opinion & Parties 23*(1): 27–48.

Green, Donald P. and Adam Zelizer. 2017. How Much GOTV Mail is Too Much? Results from a Large-Scale Field Experiment. *Journal of Experimental Political Science 4*(2): 107–18.

Kling, Daniel T., and Stratmann, Thomas. 2022. Replication Data for Large-Scale Evidence for the Effectiveness of Partisan GOTV Robo Calls. https://doi.org/10.7910/DVN/DMJ7EA. Harvard Dataverse, V1.

List, John A., Azeem M. Shaikh, and Yang Xu. 2019. Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*(4), 773–793.

Lofy, Bill. 2005. *Politics the Wellstone Way: How to Elect Progressive Candidates and Win on Issues*. Minneapolis, MN: University of Minnesota Press.

Murray, G. R. and R. E. Matland. 2014. Mobilization Effects Using Mail: Social Pressure, Descriptive Norms, and Timing. *Political Research Quarterly 67*(2): 304–19.

Nickerson, David W. 2007. Quality is Job One: Professional and Volunteer Voter Mobilization Calls. *American Journal of Political Science 51*(2): 269–82.

Nickerson, David W. 2008. Is Voting Contagious? Evidence from Two Field Experiments. *American Political Science Review 102*(01): 49–57.

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press

Nickerson, David W. and Todd Rogers. 2010. Do You Have a Voting Plan? Implementation Intentions, Voter Turnout, and Organic Plan Making. *Psychological Science 21*(2): 194–9.

Panagopoulos, Costas. 2011. Timing Is Everything? Primacy and Recency Effects in Voter Mobilization Campaigns. *Political Behavior 33*(1): 79–93.

Ramirez, Ricardo. 2005. Giving Voice to Latino Voters: A Field Experiment on the Effectiveness of a National Nonpartisan Mobilization Effort. *The Annals of the American Academy of Political and Social Science 601*(1): 66–84.

Rogers, Todd, Donald P. Green, John Ternovski, and Carolina Ferrerosa Young. 2017. Social Pressure and Voting: A Field Experiment Conducted in a High-Salience Election. *Electoral Studies 46*: 87–100.

Romano, Joseph P. and Michael Wolf. 2010. Balanced Control of Generalized Error Rates. *The Annals of Statistics 38*(1): 598–633.

Sagarin, Brad J., Stephen G. West, Alexander Ratnikov, William K. Homan, Timothy D. Ritchie, and Edward J. Hansen. 2014. Treatment Noncompliance in Randomized Experiments: Statistical Approaches and Design Issues. *Psychological Methods 19*(3): 317–33.

Shaw, Daron R., Donald P. Green, James G. Gimpel, and Alan S. Gerber. 2012. Do Robotic Calls from Credible Sources Influence Voter Turnout or Vote Choice? Evidence from a Randomized Field Experiment. *Journal of Political Marketing 11*(4): 231–45.

Zelizer, Adam. 2020. How Many Robocalls are too Many? Results from a Large-Scale Field Experiment. *Journal of Political Marketing 19*(4): 405–13.

Cite this article: Kling DT and Stratmann T (2023). Large-Scale Evidence for the Effectiveness of Partisan GOTV Robo Calls. *Journal of Experimental Political Science* **10**, 188–200. https://doi.org/10.1017/XPS.2022.16

https://doi.org/10.1017/XPS.2022.16 Published online by Cambridge University Press