

Kling, Daniel; Stratmann, Thomas

Working Paper

The Efficacy of Political Advertising: A Voter Participation Field Experiment with Multiple Robo Calls and Controls for Selection Effects

CESifo Working Paper, No. 6195

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Kling, Daniel; Stratmann, Thomas (2016) : The Efficacy of Political Advertising: A Voter Participation Field Experiment with Multiple Robo Calls and Controls for Selection Effects, CESifo Working Paper, No. 6195, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/149282>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



The Efficacy of Political Advertising: A Voter Participation Field Experiment with Multiple Robo Calls and Controls for Selection Effects

Daniel Kling
Thomas Stratmann

CESIFO WORKING PAPER NO. 6195

CATEGORY 2: PUBLIC CHOICE

NOVEMBER 2016

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

ISSN 2364-1428

CESifo

Center for Economic Studies & Ifo Institute

The Efficacy of Political Advertising: A Voter Participation Field Experiment with Multiple Robo Calls and Controls for Selection Effects

Abstract

We document the effectiveness of robo calls for increasing voter participation despite most published research finding little or no effect of automated calls. We establish this finding in a large field experiment in a targeted, partisan get-out-the-vote campaign. Our experimental design includes a follow-up call, which allows us to control for selection effects. We identify subsets of subjects, for whom the treatment effects are substantially larger than those that are found in previous studies. Our findings show that robo calls can cause a one percentage point increase in voter turnout. 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 that a few extra calls increase the treatment effect, and that many additional calls decrease that effect.

JEL-Codes: D720, C900.

Keywords: political advertising, voter turnout, selection effects, field experiment.

Daniel Kling
Department of Economics
George Mason University
4400 University Drive
USA - 22030 Fairfax VA
dkling2@gmu.edu

Thomas Stratmann
Department of Economics
George Mason University
4400 University Drive
USA - 22030 Fairfax VA
tstratma@gmu.edu

The authors thank Chris Cupit for logistical and financial support, and Donald Green, Luigi Butera, Pierre Mandon, Joy Buchanan, Angelina Christie, and seminar participants at Bocconi University and at George Mason University for helpful comments and suggestions.

I. Introduction

In light of academic research showing that automated calls do not increase voter participation, the widespread and continued use of automated get-out-the-vote (GOTV) calls in political campaigns is puzzling. That is, the overwhelming majority of field experiments show that prerecorded get-out-the-vote calls are not effective in increasing voter participation. However, these “robo calls” are still widely used by political campaigns and independent organizations. Why would actual campaigns use prerecorded GOTV calls if they are ineffective?

Robo calls, aimed at increasing voter participation in elections, are a form of advertising. Academic research has shown that the provision of information is effective in changing behavior, as for example with product advertising and political advertising.¹ From this perspective, the finding that robo calls are ineffective is surprising. Using a novel field experiment design, however, we find that partisan robo calls have a significant positive effect on turnout among registered Republican voters and their household members. In our experiment, we used separate treatment “dosages,” deploying one, three, or six robo calls in treatments T1, T3, and T6, respectively. Using a quadratic specification, we find additional calls have diminishing returns, with the largest treatment effects for subjects scheduled to receive three calls.²

Previous research shows that the impact of robo calls on voter turnout is small and not statistically significant, as summarized in a meta-study by Green, McGrath, and Aronow (2013).

¹For product advertising, please see (Akerberg, 2001). For political advertising, there is theoretical, empirical, and experimental support for its effectiveness. For example campaign expenditures influence voters by providing information about candidates, or by signaling support from well-informed donors (see Prat, 2002, and Potters et al., 1997). Altruism and group dynamics can explain how voters are mobilized when the chances of casting a pivotal vote approach zero and voting is costly (see Coate and Conlin, 2004, Feddersen, 2004, Feddersen and Sandroni, 2006, and Evren, 2012). Particularly, increasing the salience of civic duty and voter information about candidates are shown to increase turnout in structurally estimated models using election data (Degan and Merlo, 2011). When controlling for the price of television advertising, additional spending increases vote share for incumbents and challengers (Stratmann, 2009).

²In a laboratory setting, informational advertising can increase turnout (Houser et al., 2011).

They report a statistically insignificant increase in voter turnout of 0.156 percentage points, attributable to prerecorded phone messages. The magnitude of this estimate, though statistically insignificant, implies that voter participation in elections increases by one additional voter for every 641 potential voters called, though not necessarily voters reached with the placed call. Another meta-analysis reports an “average intent-to-treat effect of 0.113 percentage point, with a 95 percent 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.

An effect of robo calls on voter turnout is difficult to detect, even when analyzing data from a field experiment. One confounding factor with prerecorded calls is that the experimenter does not observe who answers a call. Another confounding factor is that subjects assigned to a treatment group may not answer any of the placed treatment calls. Thus, comparing the entire treatment group to the entire control group typically yields an intent-to-treat (ITT) effect, that is smaller than the causal treatment effect, that applies to subjects who have been successfully treated (Sagarin et al., 2014). Further, if the effect size is small, any field experiment requires a large sample to have the power to detect statistically significant effects.

The causal effect of robo calls for subjects that are successfully treated cannot be obtained by comparing only those subjects in the treatment group who answer a robo call to all subjects in the control group.³ Such a comparison introduces an upward bias, due to selection effects. A selection effect arises, because factors which lead a subject to receive treatment – because the subject answered the robo call – are not independent from factors that affect the likelihood that this subject will also participate in the election. For example, Gerber et al. (2011)

³ We define successful treatment when a subject in a household has answered in person at least one treatment call or the subject received a call on their answering machine.

find that individuals measuring high on an extraversion and emotional stability scale are more likely to vote. It is plausible that those individuals are also more likely to answer a phone call.

Gerber et al. (2010) address selection effects through an experimental design. Gerber et al. divide subjects into a control group that receives no calls, a treatment group that receives GOTV calls, and a placebo group that receives calls unrelated to voting.⁴ Studying the 2008 primary election, Gerber et al. find an approximately two percentage point increase in voter participation for those subjects who were successfully contacted, relative to subjects in the placebo group. This finding is the only academic study that reports a positive and statistically significant treatment effect of robo calls. The authors attribute the large size of their reported effect in the 2008 primary election 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 that that voting records are publicly available.⁵ However, when attempting to replicate the result found in the 2008 primary election 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 2015, p. 74).

In our study, we employ a novel method to address selection effects and we also present the first test the treatment effect of dosage on voter turnout, i.e. the effect of calling the same subject once, three, or six times over multiple days.⁶ This approach resembles how customers

⁴ See also Nickerson (2008).

⁵ Ali and Lin (2013) develop a model of voter turnout based on similar social dynamics. Using a field experiment, DellaVigna et al. (2014) present evidence that “individuals vote because they expect to be asked” about it and find lying costly.

⁶ Except for the study by Ramirez et al. (2005) who placed two separate calls, previous studies have not placed multiple calls to the same subject over several days and did therefore not test whether the treatment effect differs by

experience product advertisements, when they listen to the same advertisement on the radio or TV over several days.⁷ We estimate a statistically significant intent-to-treat effects with calls using common-place partisan get-out-the-vote messaging, as opposed to, for example, social pressure messaging.

Our dosage treatments shed additional light as to why prior studies did not find an effect of automated calls on voter participation. The treatment effect for subjects who received only a single call, was a 0.219 percentage point increase in voter participation. This estimate is larger in magnitude than previous findings, which also used single call designs, but it is not statistically significant. While the treatment effect for subjects who received six calls was also small and statistically insignificant, we find larger treatment effects among the subset of voters who were called on three days. For these voters, the treatment effect is a 0.65 percentage point increase in voter participation. This effect is statistically significant and four times larger than the average effect reported in previous meta studies, and suggests positive but diminishing marginal returns to multiple calls.⁸ This estimate does not include controls for selection effects or geographic location-based fixed effects, both of which increase the magnitude of the estimated treatment effect.

Our experimental design allows us to control for selection effects because subjects in both treatment and the control groups received a follow-up call after the election date. This

the number of calls. Shaw et al. (2012), placed one call, they designed their field experiment so that “[u]p to three attempts were made for each phone number” (Shaw et al. 2012, p. 236).

⁷ Although previous researchers find that GOTV calls are less effective the earlier these calls are placed prior to an election (Nickerson, 2007, Panagopoulos, 2011, Green and Gerber, 2015), we do not observe a timing effect across voters who received three or six calls. This suggests that within the time frame of our study, the effectiveness of our treatment does not measurably decay.

⁸ This finding is also consistent with conventional GOTV campaign practice to make “two or three contacts in the final weekend” to mobilize habitual voters, something that these campaigns consider to be most productive in generating voter turnout (Lofy, 2005, page 153).

approach allows us to estimate a voter participation effect that is closer to the causal effect of automated GOTV calls than intent-to-treat estimates. We find that the subset of treatment subjects who answered the follow-up call were also more likely to have answered the treatment call, confirming the presence of selection effects. By analyzing the behavior of only subjects who answered the post-election follow-up call, we compare a subset of treatment group subjects with a high “treatment success” contact rate to subjects in the control group with a similar propensity to answer a phone call. For this subset, the three-call treatment effect is a 1.03 percentage point increase in voter participation.

We also use geographic fixed effects to control for unobservable variation that is correlated with receiving treatment and voter participation. For example, geographic heterogeneity in weather or norms could change subject responsiveness to the treatment. These controls lead to large and statistically significant estimated treatment effects. For example, with precinct fixed effects, the estimated increase in voter participation of subjects who answered the post-election follow-up call across all treatments is 4.43 percentage points. This suggests that among certain populations, robo calls may be a potent GOTV tool, potentially explaining why political advertisers continue to deploy robo calls.

II. Theoretical Framework

We develop the intuition about why and how GOTV robo calls might be effective, and for what type of individual, from a model by Feddersen and Sandroni (2006) in which voting is costly and voters have independent preferences regarding candidates and voting. In this model, agents are of type 1 or type 2, preferring candidate 1 or 2, respectively. Candidate 1 is the candidate with minority support, such that the fraction of the electorate who prefer candidate 1 is $k \in (0, \frac{1}{2}]$.

Agents are either “ethical” – i.e., receiving a positive payoff, D , for “doing their part” by participating in the election – or “abstainers,” that is, agents who receive zero payoff for participating. Abstainers never vote. The fraction of type i agents (supporters of candidate i) who are ethical voters is $\tilde{q}_i$, which is independently and uniformly distributed over $[0,1]$. This model identifies the conditions under which an ethical agent should vote.⁹

Agents prefer an increase in the probability that their favored candidate will be elected and also have a preference to minimizing the social cost of voting (ϕ). Given these two assumptions, and defining w as the importance of an election and p as the probability that candidate 1 wins the election, an agent’s utility function is

$$U_1 = wp - \phi \text{ and } U_2 = w(1 - p) - \phi. \text{ }^{10}$$

Each agent has a cost of voting $\bar{c} \cdot z$, where $\bar{c} > 0$, and $z \sim U[(0,1)]$. For a given set of election parameters, ethical voters identify a consistent voting rule (σ_i) to determine their behavior that would optimize expected social utility if all agents adopted the appropriate rule for their type. The rule states that if a type i ethical agent’s cost of voting is higher than σ_i , then they should not vote. This rule is a function of D , k , w , and the average cost of voting ($\frac{\bar{c}}{2}$). The optimal threshold may differ between the ethical agents of type 1 and type 2, but this optimal threshold is shared by all agents of a particular type.

⁹ Ali and Lin, (2013) provide a theoretical extension, that considers social behavior and signaling.

¹⁰ Feddersen and Sandroni (2006, p. 1273) note that *ceteris paribus* “the higher the value of w , the higher the expected social cost that agents would be willing to trade for an increase in the chances that their favored candidate wins. Thus, w parameterizes the importance of the difference between the two candidates relative to the social cost of voting.”

The equilibrium conditions for this model generate predictions that are relevant to our experiment.¹¹ The model predicts that as the importance of the election (w) rises, turnout rises. Further, the model predicts that a reduced average cost of voting leads to higher turnout. A third prediction is that turnout increases as the level of disagreement in the electorate (k) increases.

While this model does not directly address efforts of campaigns or party leaders to persuade their agents to participate in elections, it suggests how GOTV treatments might affect voter turnout. As type i party leaders seek to increase turnout for their own candidates, they will attempt to change the optimal voting rule to a new $\sigma_i^{*'} > \sigma_i^*$, so that agents -- with individual voting cost ($\bar{c} \cdot z$) such that $\sigma_i^{*'} > \bar{c} \cdot z > \sigma_i^*$ -- will now choose to vote.

In the model's framework, GOTV interventions are effective if the parameters determining the optimal rule threshold are not known to agents with certainty and the robo call allows them to reduce the uncertainty regarding these parameters. Thus the robo call can provide

¹¹ A summary of the equilibrium conditions and their derivation follows. See Feddersen and Sandroni (2006) for the full model.

If ethical agents follow the rule profile (σ_1, σ_2) , then the expected social cost of voting is

$$\phi(\sigma_1, \sigma_2) \equiv \bar{c} \left(kE(\tilde{q}_1) \int_0^{\sigma_1} x \partial x + (1-k)E(\tilde{q}_2) \int_0^{\sigma_2} x \partial x \right).$$

Under this rule profile, the probability that candidate 1 is elected is given by

$$p(\sigma_1, \sigma_2) \equiv F\left(\frac{k\sigma_1}{(1-k)\sigma_2}\right),$$

where F is the cumulative distribution function of $\tilde{q}_2/\tilde{q}_1$, the ratio of the fraction of ethical agents of each type. Therefore, agents maximize

$$R_1(\sigma_1, \sigma_2) \equiv wp(\sigma_1, \sigma_2) - \phi(\sigma_1, \sigma_2); \text{ or } \\ R_2(\sigma_1, \sigma_2) \equiv w(1 - p(\sigma_1, \sigma_2)) - \phi(\sigma_1, \sigma_2).$$

The optimal rule profiles (σ_1^*, σ_2^*) are given by

$$\begin{aligned} \sigma_1^* &= \sqrt{\frac{w}{\bar{c}}} \frac{1}{\sqrt[4]{k(1-k)}}, \sigma_2^* = \sqrt{\frac{w}{\bar{c}}} \sqrt[4]{\frac{k}{(1-k)^3}} & \text{if } \frac{\bar{c}}{w} &> \frac{1}{\sqrt{k(1-k)}} \\ \sigma_1^* &= 1, \sigma_2^* = \sqrt[3]{\frac{wk}{\bar{c}(1-k)^2}} & \text{if } \frac{k}{(1-k)^2} < \frac{\bar{c}}{w} \leq \frac{1}{\sqrt{k(1-k)}} \\ \sigma_1^* &= \sigma_2^* = 1 & \text{if } \frac{\bar{c}}{w} \leq \frac{k}{(1-k)^2} \end{aligned}$$

information about the importance of the election, the cost of voting, the payoff of participating, or the level of support for the minority candidate.

Therefore, campaigns can attempt to increase turnout by changing agents' perceptions of one or more of these parameters. For example, messages focusing on the policy consequences of an election are attempts to get out the vote by increasing the perceived importance of that election, which would increase the optimal cost threshold rule. This suggests that GOTV treatments will be most effective among subjects whose costs are already close to their perceived σ_i^* . That is, a GOTV treatment will be less effective if it targets individuals who are habitual voters or individuals who never vote.¹²

This framework also suggests that treatments will be most effective among the subjects who are most receptive to messages that would change their perceptions about the election. Zaller (1992) discusses the how political engagement and the reception of information forms political opinions and responses. He suggests that those in the middle of the distribution of political knowledge and political passion are more receptive to new information. Further, Zaller suggests that messages that align to recipients' political beliefs are more likely to influence their beliefs or behavior.

Our field experiment will focus on those voters who are more likely to be closer to indifferent about voting, and who are more likely to be receptive to our treatment messages. We predict that our robo call treatments will increase voter turnout by increasing some subjects' perception of w , the importance of the election, k the level of disagreement in the population, or by decreasing their individual cost of voting by reminding them of the day and date of the election. By targeting our treatment to subjects who were more likely to be marginal voters, and

¹² See also Arceneaux and Nickerson (2009).

who were predisposed by their ideology to receive our messages, we increase the probability that our treatment will effectively increase turnout.

III. Procedure

Our field experiment occurred in the days leading up to the November 2014 general election. A political consulting firm identified registered voters in six states as subjects for our automated get-out-the-vote campaigns. 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, 2010.¹³ Within this group, we identified subjects who had voted in exactly two of the 2010 and 2012 primary and general elections and excluded all other registered voters from our subject pool.¹⁴ We also excluded subjects who voted early or who cast an absentee ballot in the 2010 or 2012 general election.

These choices were motivated by several theoretical considerations. The specific states were chosen because none of them were expected to have highly competitive statewide races. This was to increase the chance that the robo calls could influence subjects' perception of the importance of the races. We chose likely Republican voters so that subjects would be predisposed to respond to the specific political messaging in the calls. We selected subjects based on their voting history to target subjects who were more likely to be near the threshold for voting or not voting. The motivation for the early voting exclusion was that our field experiment

¹³ 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.

¹⁴ Some scholars select subjects randomly within a demographic group. One study, for example, selects subjects from a list of "registered Latino voters in low-propensity precincts" (Ramirez 2005, p. 70). Other studies select subjects that tend to be receptive to the treatment message. For example, Shaw et al. (2012, p.236), whose calls includes 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 a subject among those who voted in the past two general elections but did not vote in the most recent primary.

occurred in the days prior to the November election, and we wanted to identify subjects, who had the highest likelihood of not having cast a ballot for the November 2014 election in the weeks preceding that election.

From this group of subjects and for each state, we randomly selected 42,000 unique landline phone numbers. Many landline numbers are associated with more than one potential voter because a given landline is associated with a household consisting of more than one voter.¹⁵ Because we cannot identify which household member, if any, answers a call, we included every registered voter in a given household in the subject pool. The 42,000 phone numbers per state yielded subject pools, in each state, ranging from 86,714 to 95,557 registered voters. In total, our study includes 539,567 subjects.

Within each state, a random assignment mechanism placed subjects in one of four groups: a control group, and three treatment groups. Subjects in the three treatment groups, T1, T3, and T6 received one treatment call on one, three, and six days, respectively¹⁶. The robo calls to subjects in T6 started six days before the election, the calls to subjects in T3 started three days before the election, and all subjects in the three treatment groups received a call on the morning of Election Day. We made up to three attempts to reach the household on a given day. Once a member of the household answered the phone or an answering machine was reached, no additional attempts were made on that day. Phone numbers for multi-subject households were called the assigned number of times, regardless of the number of subjects per household. Table 1

¹⁵ Some of these household members may have already voted by mail, others household members may have a voting record that differs from having voted twice in the last four elections, and others may not be likely Republican voters.

¹⁶ Our multiple calls treatment is motivated the practice of conventional GOTV campaigns to make several contacts in the week prior to the election, with the goal to increase the likelihood that habitual voters cast a ballot on the day of the election (Lofy, 2005, page 153).

lists the number of subjects in each group by their state of residence.¹⁷ Table 2 lists the number of subjects in each group by the number of registered voters in their household. On average, each household has 2.16 members.

Each treatment call consisted of a prerecorded message with a reminder about the election date, and an encouragement to vote for Republican candidates. Messages were slightly different on each call date. Each message emphasized the importance of the election and referenced ideas that were expected to resonate with a potential Republican voter. The duration of the messages was between 35 to 45 seconds. Appendix A lists the text of each message (script). Table 3 shows the schedule of calls and schedule of scripts, which was identical for every state. Therefore, T1, T3, and T6 differed in the number of robo calls and the days those calls were placed. Further, on each day the message had a slightly changed wording.

For each call, we categorized the outcome of the call into live answer, answering machine, operator, no answer, busy, fax machine, or otherwise uncompleted.¹⁸ We define “treatment answerers” as subjects residing in households where at least one call resulted in a live answer, meaning someone in that household answered the phone. We classify all other subjects as “treatment non-answerers.”

Data for whether a subject voted in the November 2014 election come from public voting records. Our outcome variable for our analysis, that is, whether a subject participated in the election, is based on this information.

¹⁷ Once a landline is assigned to a treatment or control group, every voter in that household is assigned to the same group. The number of subjects per treatment group varies because the number of subjects per household differs.

¹⁸ 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, 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.

Households in all treatment and control groups received a follow-up call after the election. Specifically, we placed a robo call to each phone number eight months after the election. This call contained as partisan political message not related to voting in the 2014 general election. Every household in the treatment and control group received one call. As in with initial treatment calls prior to the election, we made three attempts to reach a member of the household.¹⁹ Further, and consistent with the definitions for the initial treatment, we define as follow-up answerers subjects residing in households where the follow-up call resulted in a live answer. We classify all other subjects as follow-up non-answerers.

Table 4 compares summary statistics for subject characteristics separately for the control and combined treatment groups. Although subjects were selected randomly, Table 4 shows that some subject characteristics are statistically different between both groups, though the magnitude of the difference, relative to the mean of the corresponding variables, is small. For example, subjects in the treatment groups have higher average estimated income. Further, voter participation for subjects in the control group was 0.3 percentage points higher in the 2010 general election and 0.5 percentage points lower in the 2012 general election.

Table 5 shows the summary statistics for the control and each of the three treatment groups separately. When comparing the control group to the treatment groups, there are statistically significant differences in mean age, income, and single-voter household status, though again, the magnitudes of the differences are small. In some of our regression specifications we include subject characteristic controls to address any estimation bias that might arise from unbalanced samples.

¹⁹ As the initial treatment calls, this follow-up call can includes up to three calls if the first and second call outcome was operator, no answer, busy, fax machine, or otherwise uncompleted.

IV. Results

IV. A. Baseline Results

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

Table 6 shows the intent-to-treat estimates from comparing voting rates of the three treatment groups to those of the control group. In each specification, the intercept estimate measures the percent of individuals in the control group who cast a ballot in the November 2014 general election. The coefficients on each of the treatment variables show the difference in voting rates, measured in percentage points, between the corresponding treatment group and the control group.

Table 6, Column 1 shows that the intercept estimate is 0.489, implying that 48.9 percent of subjects in the control group participated in the election. The coefficient on the *All Treatments* variable is 0.00316. These findings imply that the voter participation for subjects in the pooled three treatment group is 0.32 percentage points higher than voter participation for subjects in the control group. This point estimate is statistically significant at the ten percent level. The magnitude of this estimate implies a yield of three additional voters for every 1,000 subjects called.²⁰ This effect is nearly three times larger than the average effect reported in Green and Gerber (2015). Moreover, and contrary to previous findings, this effect is statistically different from zero. Our result implies a cost of about \$9 to induce a subject from not participating in the election to casting a ballot.²¹

²⁰ In our sample, the average number of subjects per household is 2.53. Calling 1,000 subjects implies calling approximately 400 households.

²¹ 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

Table 6, 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 three days. T3's treatment effect is 0.65 percentage points, and the corresponding point estimate is statistically significant at the one percent level.²² The estimated effect of T3 is the turnout of one additional voter for every 154 subjects called. The treatment effect for groups T1 and T6 are positive, but smaller and statistically insignificant. The finding that the T3 treatment has the largest parallels the one of the most favored approaches to generate voter turnout in conventional GOTV campaigns, that is to "two or three contacts in the final weekend" to increase turnout by habitual voters (Lofy, 2005, page 153).

Table 6, Column 2 shows that the estimates in specification Column 1 are primarily attributable to the T3 treatment. These dosage findings show that calling on three different days is more effective in increasing voter turnout than calling only on one day. Further, our findings show that calling on three days is more effective than calling on six days.²³ The Wald statistic for testing for the equality of the coefficients for T1 and T3 is 3.61 and is significant at the ten percent level. The Wald statistic testing the equality of the coefficients for T3 and T6 is 6.27 and is statistically significant at the five 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.

number of additional voters by multiplying the treatment coefficient, 0.00316, 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.

²² Following the methodology outlined in List et al. (2015), 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 C for details.

²³ Treatment T6 entailed two calls on the Saturday prior to the election and no calls on the Sunday, but for simplicity we refer to T6 as having received treatment on six days.

In households with multiple registered voters, we do not know which subject or subjects answered the treatment call or calls. Thus, in multi-voter households the treatment effect likely is diluted. Therefore, we also examine the treatment effect among single-voter households. In Table 6, Columns 3 and 4 provide those estimation results. Our data show a considerably larger treatment effect among single-voter households. In Column 3, the average treatment effect for single-voter households is 0.85 percentage points and statistically significant at the ten percent level. This estimate is more than twice the size of the corresponding estimate in Column 1.²⁴

Table 6, Column 4 separates the results by treatment group and shows the estimated effect among single-voter households. Again, the effect is positive for all three treatment groups, the T3 group has the largest effect, which is statistically significant at the five percent level. The corresponding point estimates shows that for this group, voting rates are 1.4 percentage points higher than in the control group, amounting to about twice the size of the corresponding estimate in Table 6, Column 2. The treatment effects for T1 and T6 are also about twice as large in the restricted as opposed to the unrestricted sample, though these estimated effects are not statistically significant.

While the finding that the robo call effect is larger in single voter households is consistent with the likelihood that a call to a multi-member household influences only the answerer of the call, but not other household members, this finding is also consistent with our theoretical framework. The reason for this interpretation is that we know that in single voter households, the subjects have voted in exactly two of the last four elections. Thus, these voters are predicted to be more receptive to treatment since they are likely near a marginal threshold for voting. In contrast, in addition to this type of voter, habitual voters as well as registered voters who never

²⁴ In contrast to our findings, Shaw et al. (2012) find a slightly diminished effect among single-voter households.

cast a ballot reside in multi-member households, who are more distant to the marginal threshold of voting.

Columns 5 through 8 of Table 6 show intent-to-treat estimates when controlling for subject-specific demographic and socioeconomic characteristics. Our control variables are state of residence, age, gender, educational attainment, estimated income, and number of subjects in the household.²⁵ Our estimates for these control variables are broadly consistent with the empirical literature (Smets and Van Ham (2013)). Specifically, we find that higher education, higher income, and age are correlated with higher voter participation. As in other studies, the coefficient for age-squared is negative in our sample. In contrast to prior studies, we find that males in our sample have a higher voter participation rate.

The coefficient estimate on the all treatment groups variable is 0.00326 (Table 6, Column 5). This result implies that voter participation for subjects in the treatment group is 0.326 percentage points higher than for subjects in the control group. This estimate is similar to the estimate in Column 1 of Table 6, and it is statistically significant at the ten percent level.

Table 6, Column 6 reports a treatment effect estimate for T3 of 0.00689. This result is also similar to the corresponding estimate in Column 2, and it is statistically significant at the one percent level. The estimate for T6 is 0.00429 and the magnitude of this estimate is larger than the estimate in Column 2, as well as statistically significant at the ten percent level. The estimate for T1 is negative, but statistically insignificant, as it was without including control

²⁵ 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. For the purpose of our control variables, subjects younger than 20 or older than 90 are coded as 20 or 90 years old, respectively. Subjects whose households contain more than six subjects are coded as having six subjects in the household.

variables. The estimates for subjects in single-voter households in columns 7 and 8 are similar to the estimates without controls in columns 3 and 4.

In an alternative specification to Table 6, we model the number of days on which a call was placed as a continuous dosage variable. Thus, this dosage variable takes the value of one for subjects in T1, the value of three for subjects in T3, and the value of six for subjects in the T6 treatment. The dosage variable takes the value of zero for subjects in the control group. We include the square of the dosage variable to capture diminishing returns of additional calls. We report these results in Table 7. Column 1 shows that for the entire sample, the coefficient for the dosage variable is 0.00409 and -0.000645 for the corresponding squared term. Both point estimates are statistically significant at the one percent level. Solving for the root of the first derivative suggests that voting rates will be highest with 3.2 calls. This is consistent with our consistent finding that T3 has the highest voting rates across several specifications and subsamples.

IV. B. Selection Effects

One concern about the previous estimates is that those in the treatment group who answer a robo call are not a random sample of the population. This circumstance does not allow us to easily compute the treatment on the treated effect because we do not know which subjects in the control group would have answered the phone, had they been called. Put differently, we do not observe perfect compliance in our field experiment, and since the decision not to comply is correlated with subject characteristics in the treatment group, a simple comparison of voting behavior of subjects who complied and subjects in the control group would result in biased estimates.

Our sample shows the presence of partial compliance because “only” 55 percent of the subjects in the treatment groups were in households that selected themselves into “live answerer”

treatment because someone in that household answered at least one call live. The calls to the remaining subjects in the treatment groups went to answering machines, had busy signals, or were otherwise non-connections.²⁶

Table 8 reports treatment outcomes by treatment group. For example, Table 8 shows that among subjects receiving the T3 treatment, 60 percent answered the phone live at least once during the three days that robo calls were placed to this group. Thus, 40 percent of subjects in T3 did not answer robo calls on any of the three days. Figure 1 shows a visual presentation of the last Column of the first two rows of Table 8, as well as a corresponding presentation for the control group. The control includes subjects who would have been answerers and subjects who would have been non-answerers if they had been called. However, we do not know which subjects in the control group belong to the answerers group and which subjects belong to the non-answerers group, had they been called.

Figure 2 shows the voter participation rates for treatment answerers, treatment non-answerers, and subjects in the control group. The figure shows that subjects in the control group have a higher voter participation rate than treatment non-answerers.

Figure 2 further shows that treatment compliers, i.e. answerers, have a higher propensity to vote than the general population, as represented by the control group. This suggests that answering the phone is correlated with the same traits that make a subject more likely to vote. For example, a subject may both answer a phone call and vote in the election because that

²⁶ Receiving a message on an answering machine could be considered as a form of treatment compliance. Since a treatment message left on an answering machine could induce a subject to vote, we include answering machines for the computation of IV/LATE estimates in section IV. E. However, for our discussion of selection effects, we consider answering machines and live answer to be distinct outcomes.

subject is extroverted.²⁷ Failing to account for selection effects when comparing voting rates of answerers to voting rates of all subjects in the control group upwardly biases estimates of the treatment effect.

We further investigated whether this selection might just be related to the 2014 election, or if the selection effect in the 2014 election is also informative about voting behavior in previous elections. Data for this analysis come from subjects' historical voting record. These records show that 51.7 percent of 2014 treatment compliers (live answerers) voted in the 2010 election. In contrast, 46.1 percent of 2014 treatment non-compliers voted in 2010.²⁸ This difference in past voting rates lends further support to the presence of selection effects.

IV. C. Follow-Up Calls

To address the selection effect issue, and to estimate the causal effect of robo calls on voter participation, our field experiment includes a follow-up call to all subjects both in the treatment and control groups, placed after the 2014 general election. Table 9 reports answering rates from this follow-up call, for the treatment and control groups. In both the treatment and control groups, 28 percent of subjects were follow-up answerers.²⁹ The similarity in the percentage of follow-up answerers across these groups reflects the random assignment of subjects to each of those two groups.

²⁷ Similarly, Gerber et al., (2010) find that the likelihood of not answering the phone is correlated with the likelihood of not voting.

²⁸ Even though our originally chosen subjects had voted in two of the last four elections, other members of their household, who are included in our sample, might have had a different voting record. Moreover, our originally chosen subjects might have missed the 2010 election, even though they voted in two of the last four elections. In unreported results, we find that the difference in voting rates between compliers and non-compliers is persistent since at least the 2000 general election.

²⁹ This live answer rate is lower than the live answer rate for T1 subjects when they received our automated call prior to the 2014 election, even though in both cases calls were only placed on one day. One possible explanation for the lower answering rate is the increased likelihood of relocation as well as travel since the follow-up call occurred in the summer.

Next, we examine whether the answering rates from our follow-up call provide some evidence that a follow-up call method is successful in addressing selection effects. Table 10 shows a cross tabulation for subjects in the treatment group who answered or did not answer the follow-up call, and who answered or did not answer at least one treatment call. We find that among follow-up answerers, 82 percent of subjects were also treatment answerers. Further, we find that 55 percent of follow-up non-answerers also did not answer the phone in any of the treatment calls. These results suggest that answering the follow-up call is a reasonable proxy for being inclined to answer the treatment call.

Figure 3 shows a visual presentation of the relationship between treatment and follow-up calls. That figure shows that the individuals in the treatment group, who answered our follow-up call also were more likely to answer our robo call prior to the November 2014 election.

We address the selection issue by restricting the sample to subjects in the treatment and control group who are follow-up answerers. This approach improves the comparability of control and treatment groups as shown in Table 11. This is because subjects in both sub-samples have a similar propensity to vote, measured by their likelihood of answering the follow-up call. Focusing our analysis on these subjects allows us to obtain an estimate that is closer to the causal effect of robo calls than an intent-to-treat estimate that uses the entire sample.³⁰

Figure 4 shows that voting rates for follow-up answerers in both groups are higher than for subjects who did not answer the follow up call.³¹ Moreover, voting rates of both treatment follow-up answerers and control group follow-up answers are similar, indicating that subjects in

³⁰ Our approach is an alternative to the placebo method of placing contemporaneous non-political calls to a subset of control group subjects. The goal of that method is to compare voter participation of subjects with similar propensities to answer phone calls. For a detailed description of the placebo method see Gerber et al. (2010).

³¹ See Table 12 for the full balance table between follow-up non-answerers and follow-up answerers.

both groups have a similar propensity to participate in elections. This figure also shows that among follow-up answerers and follow-up non-answerers, treatment subjects have slightly higher voting rates than control subjects, foreshadowing the finding that robo calls are effective in increasing voter turnout. Next, we test whether this difference in voter participation shown in Figure 4 is attributable to the fact that the subjects in the treatment groups received a robo call.

IV. D. ITT Estimation Results, Controlling for Selection Effects

Table 13 reports intent-to-treat effects from the subset of follow-up answerers in the treatment and control groups.³² Columns 1 and 2 report results for subjects who answered the follow-up with a live answer. Columns 3 and 4 report results for subjects whose follow-up call went to an answering machine. For these subjects, we predict that these coefficients are smaller in magnitude than those in Columns 1 and 2, since for those individuals, some of the selection effect might still be present. Columns 5 and 6 report results for subjects whose follow-up call was neither answered live nor went to an answering machine. For these specifications, we predict that the presence of the selection effect in this subsample will also generate smaller point estimates on the treatment variables than in Columns 1 and 2.

Compared to results from the unrestricted sample in Table 6, both the base voting rate and the average treatment effect are larger among the subset of follow-up answerers. The coefficient for live answer follow-ups for all treatments is 0.00553, implying a more than half a percentage point increase in the probability of voter participation of a household member, when a subject in that household is a member of the treatment group. This estimate is statistically significant at the eleven percent level, and the magnitude of this estimate is five times larger than

³² The results in Table 13 measure intent-to-treat effects because 18 percent of follow-up answerers were treatment non-answerers.

in a meta study by Green et al (2013). Another way to interpret the coefficient is that this estimate implies an expected six additional voters participating in the 2014 election for every 1,000 subjects called.

Column 2 reports results by treatment. The treatment effects for T1 and T6 are both positive and larger than the corresponding estimates in Table 6, while statistically insignificant. The coefficient for T3 is 0.0103 and statistically significant at the five percent level.³³ The coefficients are jointly statistically significant at the ten percent level. The estimate for T3 implies that voter participation for live follow-up answerers in the T3 group is 1.03 percentage points higher than for live follow-up answerers in the control group. The magnitude of this estimate implies an additional ten individuals participating in an election for every 1,000 subjects called. Given that the T3 group includes 37,112 follow-up answerers, the 0.0103 point estimate on T3 shows that robo calls caused at least an additional 371 subjects to participate in the 2014 election.³⁴

In Table 13, Columns 3 and 4 report the treatment effects for those subjects whose follow-up call went to an answering machine. These coefficients are all smaller and statistically insignificantly different from the control group. Similarly, Columns 5 and 6 show that for those who did not receive the follow-up call either live or via an answering machine, the point estimates on the treatment variables are small and are statistically insignificant, except for T3,

³³ After correcting for multiple hypothesis testing with three subgroups and three treatment groups, we find that this estimate remains statistically significant at the five percent level. See Appendix C.

³⁴ As in Table 6, we estimated treatment effects for subjects in single-voter households who were follow-up answerers. However, this subsample contained fewer than 14,000 subjects. In this restricted sample, the coefficient for all treatments is 0.0100, but is not statistically significant. The coefficient for T3 among follow-up answerers in single-voter households is 0.025 and is statistically significant at the five percent level.

where the size of the coefficient is 0.0074 and is statistically significant at the ten percent level.³⁵

The estimates in these columns are not jointly significant.

Table 14 reports intent-to-treat effects from the subset of follow-up answerers in the treatment and control groups, after including demographic and socioeconomic characteristics in the regression model. As in previous specifications, we use subject state, age, gender, educational attainment, estimated income, and number of subjects as control variables. Several estimates which were not significant in Table 13 become marginally significant with the introduction of control variables; however, this does not appreciably change most estimates of the treatment effects. This result indicates that restricting the sample to follow-up answerers does not disrupt the random assignment of subjects to treatment and control groups.

IV. E. Local Average Treatment Effects

To verify that that restricting the sample to follow up answerers allows us to get a better estimate of the causal effect, we compute the Local Average Treatment Effect (LATE) (Angrist and Pischke, 2014). The LATE is a two-stage least squares (2SLS) IV estimate of the average causal effect for subjects who were treatment answerers. That is, calculating the LATE allows us to estimate the magnitude of the treatment effect on the subjects who received the treatment.³⁶ We use the random assignment to the treatment group as an instrument for answering a treatment call live or by answering machine. A valid instrument satisfies three criteria. One is that the instrument must have a causal effect on the causal variable of interest. Second the instrument

³⁵ After correcting for multiple hypothesis testing with three subgroups and three treatment groups, we find that this estimate is not statistically significant at the ten percent level.

³⁶ This estimate may also be of more interest to GOTV practitioners, as it is common for campaigns to pay only for calls that are answered by a person or by an answering machine.

must be randomly assigned and be unrelated to important omitted variables. Third, the instrument must only affect the dependent variable through the causal variable of interest

Our application meets these three criteria. Therefore the LATE calculation is an estimate of the causal treatment effect on subjects who were treatment answerers or received a message on their answering machine.

We find that estimated LATE for all subjects is 0.42 percentage points, and that this estimate is statistically significant at the five percent level (Table 15, Column1). This magnitude implies an expected four additional voters for every 1,000 subjects reached. Table 15, Columns 2 through 4 report the LATE estimates for each treatment group, and Columns 5 through 8 report the corresponding LATE estimates for follow-up answerers.³⁷ Column 3 shows that the LATE for T3 is 0.86 percentage points and Column 7 shows that the LATE for T3 among only follow-up answerers is 1.1 percentage points. Both of these estimates are statistically significant at the one percent level. The latter estimate implies an expected eleven additional voters for every 1,000 subjects reached in T3. The estimates for T1 and T6 are not statistically significant. These estimates support our previous findings that the T3 treatment generates the largest treatment effect and the robo calls are an effective tool to increase voter participation.

IV. F. Dosage and Call Timing

In earlier tables, we established that T3 is the most effective treatment. However, within T3 or T6, subjects may have answered different numbers of calls, on different days, and heard different scripts. To analyze the effect of dosage in more detail, Table 16 presents results from regression models that test the effect of the number of successful treatment calls that were answered live in

³⁷ We compute the LATE for the sub-sample of follow-up answerers since this group does not have full compliance, though compliance rates are higher than the overall sample. Specifically, for follow up answers compliance is 82 percent (Table 10).

a household on whether a subject in that household voted in the 2014 general election.³⁸

Specifically, Table 16 shows the expected difference in voting rates from the control group for subjects in T3 and T6. This expected difference is based on how many calls to a subject's phone number were answered live.

The point estimates in Table 16 capture both the selection effects associated with compliance, as well as the effect of receiving different numbers of robo calls.³⁹ Table 16, Column 1 shows that the turnout rate for subjects in T3 with no live answers is 4.7 percentage points lower than for subjects in the control group. The voting rates for subjects with one, two, or three live answers are 3.2, 4.3, and 5.1 percentage points higher than for subjects in the control group, respectively. All four estimates are statistically significant at the one percent level. Column 3 shows that the turnout rate for subjects in T6 with no live answers is 6.3 percentage points lower than for subjects in the control group. The estimates for one, two, three, and four live answers also are positive and increase monotonically. However, the turnout rates for five and six live answers are slightly lower than for four answers, though none of these three estimates is statistically different from the others.

Columns 2 and 4 restrict the sample to follow-up answerers, which mitigates selection effects. Column 2 shows that the turnout rate for subjects in T3 with no live answers is 2.7 percentage points lower than for subjects in the control group. The turnout rate for subjects with one live answer is almost the same as the control. The turnout rate for subjects with two or three

³⁸ Detailed subject call outcome data, including the number of subjects or households associated with each call outcome per treatment group, is provided in Appendix B.

³⁹ For example, a subject in T3 who answered the last two calls would have heard the same scripts as a subject in T6 who only answered the last two calls. However, the T3 subject would have answered two of three attempted calls, while the T6 subject answered one-third of the attempted calls.

live answers is almost two percentage points higher than for subjects in the control group, and the difference between two and three live answers is small.

Column 4 shows a similar pattern for subjects in T6. Turnout for subjects in T6 with two or fewer live answers is lower than turnout for control group subjects. Turnout for subjects with four or five live answers is approximately 1.6 percentage points higher than for subjects in the control and for subjects with six live answers, the turnout rate is only 1.3 percentage points higher. These results show a diminishing, or possibly negative, effect for a high number of robo calls. Although these results are suggestive, with these data, we cannot decompose selection effects and the effect of received treatment dosage.

Even subjects who answered the same number of calls may have answered calls earlier or later in the treatment. This affects the timing of the treatment received, as well as the scripts heard, since each day's script was somewhat different (see Appendix A). Previous research has examined whether the timing of GOTV interventions affects their effectiveness. Nickerson (2007) finds that GOTV phone calls delivered the week prior to an election are more effective than earlier calls. Panagopoulos (2011) compares phone treatments from four weeks, two weeks, and three days prior to an election and does not find significant differences between effect sizes.

Although all of the calls in our treatment were placed in the week before the election, we analyzed whether the timing of the calls answered by subjects predicted voter participation. That analysis, unreported here, shows that the day a subject receives a treatment call does not affect voting rates. This suggests that within the week prior to an election, call timing may not be an important factor. It also suggests that the different scripts did not generate quantitatively important and statistically significant different treatment effects.

IV. G. Geographic Fixed Effects

Voter participation in an election may be affected by local factors such as the socio-economic make-up of the precinct, local weather, local GOTV and campaign efforts, etc. Some of these factors may also influence the effectiveness of our treatment calls. For example, treatment calls might be correlated with areas that have saturated GOTV coverage. At the same time this GOTV coverage might lead to high voter participation. In this case, OLS underestimates the effect of treatment calls. In particular, given that our sample includes likely Republican voters, areas where many of these voters reside might have high GOTV coverage and are more than proportionally covered in our treatment.

Applying location fixed effects allows us to control for geographic characteristics, such as the propensity of voters in a given neighborhood, to vote for a Republican candidate. Specifically, geographic fixed effects will also capture background variables that lead both to an increase or decrease in the likelihood to answer a phone and the likelihood to vote, thus capturing the similarity of individuals within a geographic area.

We estimate three sets of fixed effects regressions. One set includes city fixed effects, another set includes county fixed effects and a third set includes precinct fixed effects. Table 17 through Table 19 present the results that replicate Columns 1 and 2 from Table 13 and Table 14, and include the aforementioned fixed effects. In these specifications reported in Tables 15, 16, and 17, we restrict our sample to the subset of subjects who answered the follow-up call.

Table 17 Column 1 shows that with county fixed effect controls, the coefficient on the all treatments variable implies a 1.5 percentage points increase in voter participation. This estimate is statistically significant at the five percent level. With city fixed effect controls, the coefficient for all treatments increases 1.7 percentage points and is also statistically significant at the five

percent level. With precinct fixed effect controls, the coefficient for all treatments is 4.4 percentage points, statistically significant at the five percent level. These effects are considerably larger than any of the robo call treatment effects reported in the literature.

The treatment effects separated by treatment group are quite large, and differ from the pattern reported in previous specifications in that in the fixed effects specifications, most treatments generate statistically significant effects and have a larger effect on voter participation. With county fixed effects, the coefficient for T3 is 1.5 percentage points. This estimate is statistically significant at the five percent level. The coefficients for T1 and T6 are similar and statistically significant at the ten percent level. With city fixed effects, the coefficient for T3 is 1.8 percentage points. This estimate is statistically significant at the five percent level. The coefficient for T1 is slightly larger and significant at the ten percent level, while the estimate for T6 is close to one percentage point, though not statistically significant. With precinct fixed effects, the coefficients increase with the number of calls per group, and all of the estimates are statistically significant.

Previous sections of our paper show that our empirical design generates ITT estimates and IV estimates that, contrary to much of the previous literature are statistically significant, and quantitatively important. Our follow-up design allows more precise comparison of treated subjects to the relevant control group. These effects become even larger when introducing geographic fixed effects. Our results suggest that local characteristics play an important role in determining the effectiveness of robo calls. Moreover, these results show why robo calls continue to be a staple of political campaigns.

V. Conclusion

Prior to this study, the only robo call study to find a significant treatment effect used a social pressure message (Gerber et al., 2010). Green and Gerber (2015) attribute this study's success to the effectiveness of the messaging.

Our results show that targeted robo calls have a positive effect on voter turnout even without using a social pressure message. We find that, on average, the intent-to-treat effect for T3 is four to six times larger than previous measures of robo call effectiveness. 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 has the smallest effect. These results suggest that additional calls can increase effectiveness, but that too many calls may be counterproductive. The specification with precinct fixed-effects is the one exception, showing the strongest effect in the treatment with six calls, though the treatment group point estimates are not statistically different from each other. This may warrant further experiments, using placebos, to determine the causal effect of receiving multiple calls.

We apply an experimental follow-up method to reduce the selection biases inherent to an experiment where subjects opt in to treatment. We find the largest treatment effects among subjects who answer a follow-up call, particularly subjects in T3. This effect is statistically significant. When we use this follow-up method and apply location fixed effects, the treatment effect size is comparable to the effects found in treatments with live volunteer phone banks which are shown to raise voter participation in elections by 2.9 percentage points (Green and Gerber, 2015 page 196).

In most specifications, treatment effects are larger for samples of single-voter households than for the entire sample, which includes households with more than one voter. Because of the

relatively small number of subjects in single-voter households, not all of these estimates are statistically significant. However, the pattern of greater measured treatment effect size in single-voter households is consistent.

Several explanations are consistent with a larger treatment effect for single-voter households. This finding is consistent with our theoretical framework, because subjects in single-voter households are members of the originally selected subject pool, and are predicted to be more likely to respond to the treatment. However, an alternative interpretation of this result is that the estimate for multi member households is diluted. That is, even in households with live answers, not every voter in that household is necessarily exposed to the treatment message. Another explanation for this finding comes from the observation that baseline voting rates are lower for single-voter households, and that therefore marginal effect is larger for rob calls for increasing turnout. For future research, it is of interest to disentangle these explanations.

A theoretical framework predicting that GOTV treatments are most effective with for potential voters who are more marginal voters and whose ideology predisposes them be receptive such messages, informed the design of our experiment. That theory guided this design, may explain why we find robust and statistically significant treatment effects, while broadly targeted experiments have tended to find null results.

In this experiment we employ multiple-call dosage and thus our treatments mimic political campaigning as it is practiced. Our results suggest that it is not irrational for campaigns to deploy robo calls as a cheap and additional tool to increase voter turnout. We find that our treatment increases voter participation by between three to ten additional voters for every 1,000 subjects called. This corresponds to a cost of \$7 to \$9 to induce a subject to vote in elections.

References

- Ackerman, D. A. (2001). Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, 316-333.
- Ali, S. N., & Lin, C. (2013). Why People Vote: Ethical Motives and Social Incentives. *American Economic Journal: Microeconomics*, 5(2), 73-98.
- Angrist, J. D., & Pischke, J. S. (2014). Mastering 'metrics: The path from cause to effect. Princeton University Press.
- 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.
- Coate, S., & Conlin, M. (2004). A group rule–utilitarian approach to voter turnout: theory and evidence. *The American Economic Review*, 94(5), 1476-1504.
- Degan, A., & Merlo, A. (2011). A structural model of turnout and voting in multiple elections. *Journal of the European Economic Association*, 9(2), 209-245.
- DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2014). Voting to tell others (No. w19832). National Bureau of Economic Research.
- Evren, Ö. (2012). Altruism and voting: A large-turnout result that does not rely on civic duty or cooperative behavior. *Journal of Economic Theory*, 147(6), 2124-2157.
- Feddersen, T. J. (2004). Rational choice theory and the paradox of not voting. *The Journal of Economic Perspectives*, 18(1), 99-112.
- Feddersen, T., & Sandroni, A. (2006). A theory of participation in elections. *The American Economic Review*, 96(4), 1271-1282.
- Gerber, A. S., Green, D. P., Kaplan, E. H., & Kern, H. L. (2010). Baseline, placebo, and treatment: Efficient estimation for three-group experiments. *Political Analysis*, 18(3), 297-315.
- Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., Raso, C., & Ha, S. E. (2011). Personality traits and participation in political processes. *The Journal of Politics*, 73(03), 692-706.
- Green, D. P., & Gerber, A. S. (2015). Get out the vote: How to increase voter turnout, 3rd ed. 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 & Parties*, 23(1), 27-48.

Houser, D., Morton, R., & Stratmann, T. (2011). Turned on or turned out? Campaign advertising, information and voting. *European Journal of Political Economy*, 27(4), 708-727.

List, J. A., Shaikh, A. M., & Xu, Y. (2016). Multiple hypothesis testing in experimental economics (No. w21875). National Bureau of Economic Research.

Lofy, B. (2005). *Politics the Wellstone way: How to elect progressive candidates and win on issues*. Minneapolis, MN: University of Minnesota Press.

Nickerson, D. W. (2007). Quality is job one: Professional and volunteer voter mobilization calls. *American Journal of Political Science*, 51(2), 269-282.

Nickerson, D. W. (2008). Is voting contagious? Evidence from two field experiments. *American Political Science Review*, 102(01), 49-57.

Panagopoulos, C. (2011). Timing is everything? Primacy and recency effects in voter mobilization campaigns. *Political Behavior*, 33(1), 79-93.

Potters, J., Sloof, R., and Van Winden, F. (1997). Campaign expenditures, contributions and direct endorsements: The strategic use of information and money to influence voter behavior. *European Journal of Political Economy* 13(1), 1-31.

Prat, A. (2002). Campaign spending with office-seeking politicians, rational voters, and multiple lobbies. *Journal of Economic Theory*, 103(1), 162-189.

Ramirez, R. (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.

Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological Methods*, 19(3), 317.

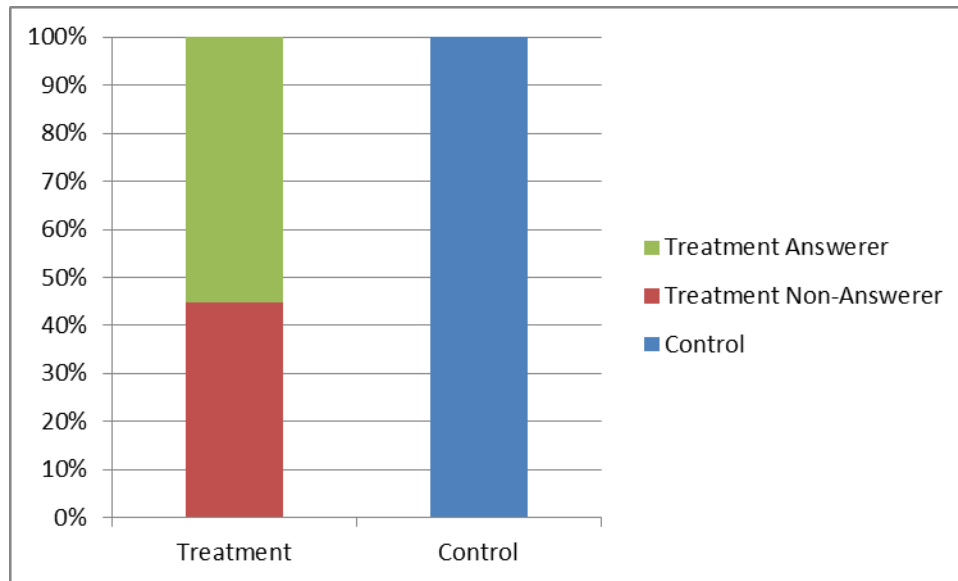
Shaw, D. R., Green, D. P., Gimpel, J. G., & Gerber, A. S. (2012). Do robotic calls from credible sources influence voter turnout or vote choice? Evidence from a randomized field experiment. *Journal of Political Marketing*, 11(4), 231-245.

Smets, K., & Van Ham, C. (2013). The embarrassment of riches? A meta-analysis of individual-level research on voter turnout. *Electoral Studies*, 32(2), 344-359.

Stratmann, T. (2009). How prices matter in politics: the returns to campaign advertising. *Public Choice*, 140(3-4), 357-377.

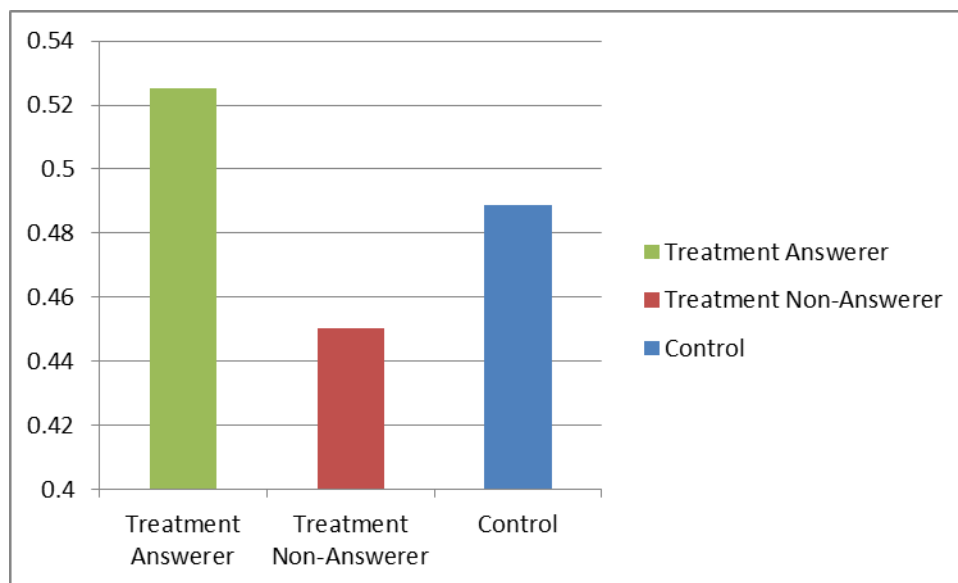
Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Figure 1: Treatment answer rates



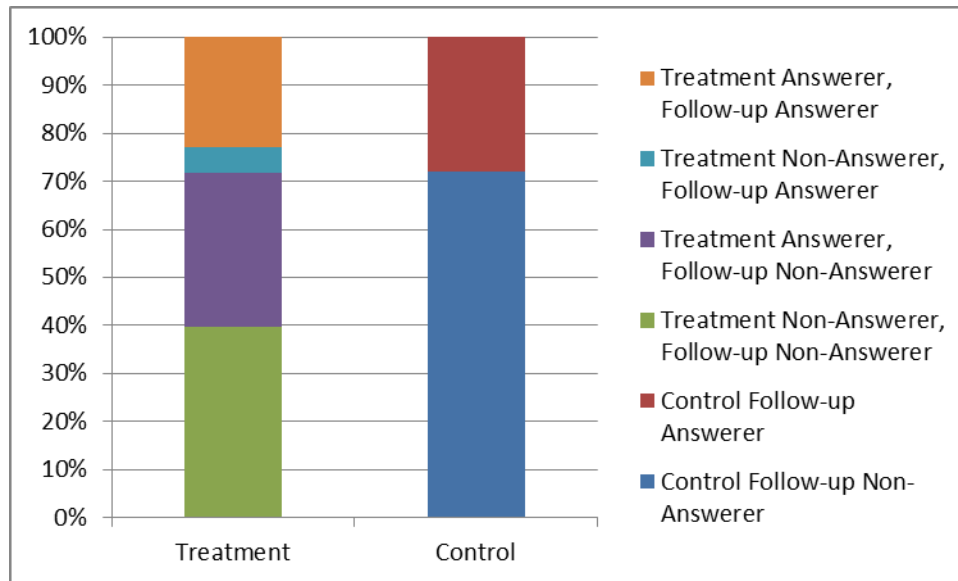
Note: Column height indicates the percentage of subjects in a category. Subjects in the treatment group are divided into subjects who answered a treatment call, and subjects who did not answer a treatment call. Subjects in the control group were not called, so we cannot distinguish which subjects in the control group would have answered a treatment call.

Figure 2: Voting rates by treatment outcome and subject group



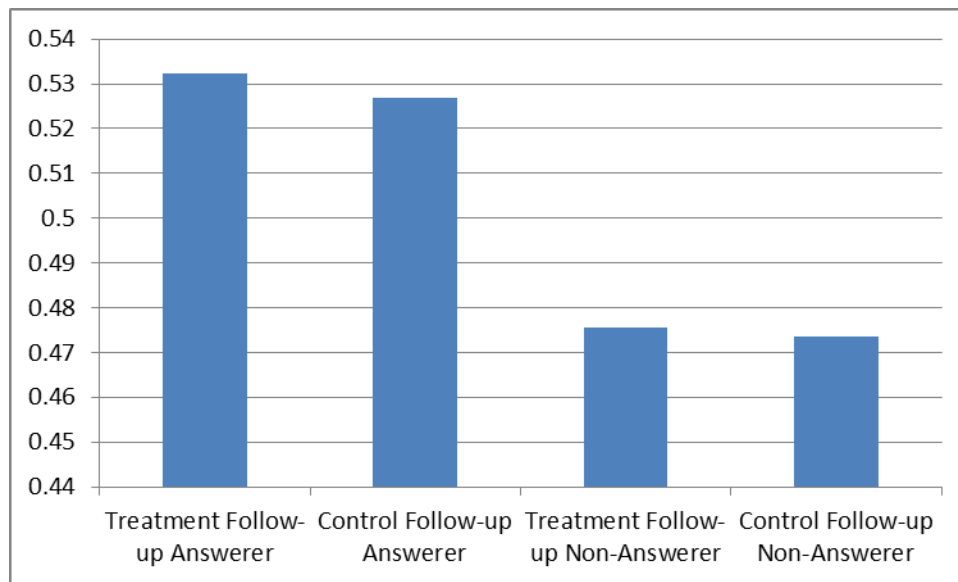
Note: Column height indicates the percentage of subjects within each category who voted in the 2014 general election.

Figure 3: Treatment success rates by follow up call response



Note: Column height indicates the percentage of subjects in a category. Subjects in the treatment group are categorized based on whether or not they answered a treatment call, and whether or not they answered the follow up-call. Subjects in the control group are categorized based on whether or not they answered the follow-up call.

Figure 4: Voting rates by follow-up call outcome and subject group



Note: Column height indicates the percentage of subjects within each category who voted in the 2014 general election. Treatment answerers and non-answerers are combined into groups of follow-up answerers and follow-up non-answerers.

Table 1: Number of subjects by state and treatment group

State	Control	T1	T3	T6	Total
Georgia	21,571	21,937	21,807	21,399	86,714
Nebraska	22,464	22,202	22,663	22,749	90,078
New Mexico	23,284	22,985	22,111	22,166	90,546
Ohio	23,030	22,804	22,681	22,882	91,397
Pennsylvania	21,356	21,534	21,541	20,844	85,275
Virginia	23,626	24,458	24,023	23,450	95,557
Total	135,331	135,920	134,826	133,490	539,567

Note: Each cell includes the number of subjects in the six states used in our study. The first Column shows the number of subjects in the control group by state. The Column labeled T1 shows the number of subjects in households that were assigned to be called one time. Column T3 shows the number of subjects in households assigned to be called three times. Column T6 shows the number of subjects in households assigned to be called six times.

Table 2: Number of subjects by household size

Registered voters per household	Control	T1	T3	T6	Total
1	13,740	13,470	13,399	14,240	54,849
2	61,434	61,616	62,598	61,294	246,942
3	40,029	39,207	39,369	39,405	158,010
4	15,096	16,004	14,676	14,064	59,840
5	4,030	4,385	3,900	3,565	15,880
6	780	948	684	696	3,108
7	147	231	126	210	714
8	48	40	56	16	160
9	27	9	18	0	54
10	0	10	0	0	10
Total	135,331	135,920	134,826	133,490	539,567

Note: This table shows how many subjects assigned to each treatment group are members of households by household size.

Table 3: 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 A includes the text for each of the six scripts.

Table 4: Balance table: Subject characteristics comparing all treatment groups to control

	Control	All treatments	Difference	Prob > T
Mean Age	49.9	50.0	-0.1	0.114
	(0.05)	(0.03)	(0.05)	
Mean Income	72,185	73,591	-1,406	< 0.001
	(130.7)	(79.9)	(157.2)	
Percent Male	49.3%	49.4%	-0.2%	0.319
	(0.1)	(0.1)	(0.2)	
Percent in SVH	10.2%	10.2%	0.0%	0.861
	(0.1)	(0.0)	(0.1)	
Percent voted in 2010	49.5%	49.2%	0.3%	0.089
	(0.1)	(0.1)	(0.2)	
Percent voted in 2012	76.7%	77.2%	-0.5%	< 0.001
	(0.1)	(0.1)	(0.1)	
Observations	135,331	404,236	539,567	

Note: This table shows average subject characteristics for the control and combined treatment groups. Percent in SVH is the percent of subjects in single-voter households. Column 3 shows the difference between the control and treatment averages for a given characteristic. Column 4 shows the p-value from the t-test for the difference in means. Standard errors are displayed in parentheses.

Table 5: Balance table: Subject characteristics by treatment group and differences from the control

	T1	T1 vs. control difference	T3	T3 vs. control difference	T6	T6 vs. control difference
Mean Age	49.7	0.2	49.9	0.0	50.3	-0.4
	(0.05)	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)
Mean Income	79,710	-7,526	74,353	-2,169	66,609	5,576
	(149.0)	(198.3)	(140.0)	(191.6)	(121.5)	(178.6)
Percent Male	49.7%	-0.4%	49.3%	0.0%	49.3%	-0.1%
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)	(0.2)
Percent in SVH	9.9%	-0.4%	9.9%	0.3%	10.7%	-0.6%
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)	(0.2)
Percent voted in 2010	49.1%	0.3%	49.3%	0.2%	49.2%	0.3%
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)	(0.2)
Percent voted in 2012	77.1%	-0.5%	77.1%	-0.4%	77.2%	-0.6%
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)	(0.2)
Observations	135,920		134,826		133,490	

Note: This table shows average subject characteristics for the control and each treatment group. Percent in SVH is the percent of subjects in single-voter households. Columns 2, 4, and 6 show the treatment group averages subtracted from the control group average. Standard errors are displayed in parentheses. Thus, the ratio of the difference in characteristics between treatment and controls and the standard error, provides the t-test for the difference in means.

Table 6: Intent-to-treat effect for all subjects and single-voter households (SVH)

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted	(5) voted	(6) voted	(7) voted	(8) voted
Population	All subjects	All subjects	SVH	SVH	All subjects	All subjects	SVH	SVH
T1		0.00219 (0.00225)		0.00755 (0.00590)		-0.00137 (0.00223)		0.00589 (0.00597)
T3		0.00647*** (0.00226)		0.0136** (0.00591)		0.00689*** (0.00224)		0.0151** (0.00599)
T6		0.000798 (0.00226)		0.00442 (0.00582)		0.00429* (0.00225)		0.00409 (0.00591)
All Treatments	0.00316* (0.00184)		0.00845* (0.00479)		0.00326* (0.00182)		0.00826* (0.00485)	
Constant	0.489*** (0.00159)	0.489*** (0.00159)	0.380*** (0.00414)	0.380*** (0.00414)				
Subject controls	No	No	No	No	Yes	Yes	Yes	Yes
Observations	539,567	539,567	54,849	54,849	515,653	515,653	51,402	51,402
R-squared	7.510e-06	2.495e-05	5.648e-05	1.028e-04	0.114	0.114	0.054	0.054
F-test	2.951	3.253	3.111	1.876	2574	2397	130.6	121.4
Prob > F	0.0858	0.0207	0.0778	0.131	< 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 voter age, gender, education, income, number of subjects in the household, and state of residence. Data for control variables was not available for a small number of subjects leading to slightly fewer observations in the columns with controls than in the Columns without controls. Standard errors are clustered by household and displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Intent-to-treat effect based on the number of attempted calls for all subjects and single-voter households with demographic and socioeconomic controls

VARIABLES	(1) voted	(2) voted	(3) Voted	(4) voted
Population	All subjects	SVH	All subjects	SVH
Call Days	0.00409*** (0.00115)	0.00838** (0.00353)	0.00355*** (0.00134)	0.00925*** (0.00359)
Days squared	-0.000645*** (0.000181)	-0.00128** (0.000555)	-0.000419** (0.000211)	-0.00141** (0.000564)
Constant	0.488*** (0.00122)	0.380*** (0.00375)		
Subject controls	No	No	Yes	Yes
Observations	539,567	54,849	515,653	51,402
R-squared	2.388e-05	1.026e-04	0.114	0.054
F-test	6.443	2.814	2483	125.9
Prob > F	0.00159	0.0599	< 0.001	< 0.001

Note: The table shows OLS estimates for the expected increase in voting rates based on the number of days that a subject received a call: zero, one, three, or six. The number of days squared is included to model diminishing returns. The subset of subjects used to estimate each specification is described in the population row. Subject controls include voter age, gender, education, income, number of subjects in the household, and state of residence. Data for control variables was not available for a small number of subjects leading to slightly fewer observations in Columns 3 and 4. Standard errors are clustered by household and displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Treatment call outcomes answer by treatment group

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

Table 9: Follow-up answering rates in control and treatment groups

Follow-up result	Control	T1	T3	T6	Total
Live answer	28% (37,622)	28% (38,282)	28% (37,112)	28% (37,948)	28% (150,964)
Answering machine	41% (55,900)	42% (57,314)	41% (55,212)	40% (53,320)	41% (221,746)
No answer	31% (41,809)	30% (40,324)	32% (42,502)	32% (42,222)	31% (166,857)
Total	100% (135,331)	100% (135,920)	100% (134,826)	100% (133,490)	100% (539,567)

Note: The table shows the percent of subjects in each group for whom the follow-up call to their household resulted in a live answer, an answering machine responding, or in no answer. These follow-up call outcome categories are listed in the rows. Percents are relative to the total number of subjects in each treatment group. The raw number of subjects in each cell is given in parentheses.

Table 10: Cross tabulation between treatment answerers and follow-up answerers by treatment group

	T1		T3		T6	
	Treatment answerer	Treatment non-answerer	Treatment answerer	Treatment non-answerer	Treatment answerer	Treatment non-answerer
Follow-up answerer	63% (24,033)	37% (14,249)	89% (32,904)	11% (4,208)	94% (35,673)	6% (2,275)
Follow-up non-answerer	30% (29,322)	70% (68,316)	49% (47,393)	51% (50,321)	56% (53,959)	44% (41,583)
Total	39% (53,355)	61% (82,565)	60% (80,297)	40% (54,529)	67% (89,632)	33% (43,858)

Note: The table shows the percent of subjects in each treatment group who were treatment answerers or treatment non-answerers for follow-up answerers and follow-up non-answerers. Percents are relative to the total number of subjects in each treatment group conditional on the follow-up call outcome.. The raw number of subjects in each cell is given in parentheses.

Table 11: Balance table: Subject characteristics comparing all treatment groups to control for follow-up answerers only

	Control	All treatments	Difference	Prob > T
Mean Age	52.7 (0.09)	52.7 (0.05)	-0.01 (0.11)	0.933
Mean Income	70,100 (245.1)	72,409 (151.8)	-2,309 (298.6)	< 0.001
Percent Male	49.3% (0.3)	49.2% (0.2)	0.1% (0.3)	0.796
Percent in SVH	8.9% (0.1)	9.0% (0.1)	0.0% (0.2)	0.793
Percent voted in 2010	52.1% (0.3)	51.7% (0.1)	0.3% (0.3)	0.247
Percent voted in 2012	78.0% (0.2)	78.4% (0.1)	-0.4% (0.2)	0.068
Observations	37,622	113,342	150,964	

Note: This table shows average subject characteristics for the control and combined treatment groups among subjects in households that answered the post-election follow-up call. Percent in SVH is the percent of subjects in single-voter households. Column 3 shows the difference between the control and treatment averages for a given characteristic. Column 4 shows the p-value from the t-test for the difference in means. Standard errors are displayed in parentheses.

Table 12: Balance table: Subject characteristics comparing follow-up answerers to follow-up non-answerers.

	Follow-up non-answerers	Follow-up answerers	Difference	Prob > T
Mean Age	48.9	52.7	-3.9	< 0.001
	(0.03)	(0.05)	(0.05)	
Mean Income	73,793	71,832	1,961	< 0.001
	(80.3)	(129.3)	(151.4)	
Percent Male	49.5%	49.3%	0.2%	0.219
	(0.1)	(0.1)	(0.2)	
Percent in SVH	10.6%	9.0%	1.7%	< 0.001
	(< 0.1)	(0.1)	(0.1)	
Percent voted in 2010	48.3%	51.8%	-3.5%	< 0.001
	(0.1)	(0.1)	(0.2)	
Percent voted in 2012	76.5%	78.3%	-1.8%	< 0.001
	(0.1)	(0.1)	(0.1)	
Observations	388,603	150,964	539,567	

Note: This table shows average subject characteristics for follow-up non-answerers and follow-up answerers in the entire subject pool. Percent in SVH is the percent of subjects in single-voter households. Column 3 shows the difference between the follow-up non-answerers and follow-up answerers averages for a given characteristic. Column 4 shows the p-value from the t-test for the difference in means. Standard errors are displayed in parentheses.

Table 13: Intent-to-treat effect among subjects by follow up call outcome

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted	(5) voted	(6) voted
Population	Follow-up answerer	Follow-up answerer	Follow-up AM	Follow-up AM	Follow-up non-answerer	Follow-up non-answerer
T1		0.00534 (0.00419)		-0.00318 (0.00345)		0.00237 (0.00408)
T3		0.0103** (0.00424)		0.00488 (0.00348)		0.00738* (0.00405)
T6		0.00103 (0.00420)		-0.000844 (0.00351)		0.00498 (0.00405)
All treatments	0.00553 (0.00344)		0.000253 (0.00284)		0.00495 (0.00331)	
Constant	0.527*** (0.00298)	0.527*** (0.00298)	0.519*** (0.00245)	0.519*** (0.00245)	0.413*** (0.00287)	0.413*** (0.00287)
Subject controls	No	No	No	No	No	No
Observations	150,964	150,964	221,746	221,746	166,857	166,857
R-squared	2.293e-05	6.594e-05	4.826e-08	3.477e-05	1.896e-05	3.177e-05
F-test	2.580	2.458	0.00793	1.904	2.238	1.244
Prob > F	0.108	0.0609	0.929	0.127	0.135	0.292

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 intercept represents the baseline voting rate in the control group. The subset of subjects used to estimate each specification is described in the population row. Standard errors are clustered by household and displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 14: Intent-to-treat effect among subjects who answered the follow up call with demographic and socioeconomic controls

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted	(5) voted	(6) voted
Population	Follow-up answerer	Follow-up answerer	Follow-up AM	Follow-up AM	Follow-up non-answerer	Follow-up non-answerer
T1		0.00351 (0.00417)		-0.00706** (0.00342)		0.000818 (0.00411)
T3		0.00963** (0.00420)		0.00581* (0.00345)		0.00837** (0.00408)
T6		0.00577 (0.00417)		0.00201 (0.00348)		0.00674* (0.00409)
All treatments	0.00628* (0.00341)		0.000136 (0.00281)		0.00539 (0.00334)	
Subject controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	146,037	146,037	213,941	213,941	155,675	155,675
R-squared	0.117	0.117	0.121	0.121	0.093	0.093
F-test	743.6	692.3	1143	1065	612.8	570.7
Prob > F	< 0.001	< 0.001	< 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 described in the population row. Subject controls include voter age, gender, education, income, number of subjects in the household, and state. Data for control variables was not available for a small number of subjects leading to slightly fewer observations than in Table 13. Standard errors are clustered by household and displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 15: Local Average Treatment Effect (LATE) estimates

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted	(5) voted	(6) voted	(7) voted	(8) voted
Population	All subjects	T1, control	T3, control	T6, control	Follow-up answerers	T1, control, follow-up answerers	T3, control, follow-up answerers	T6, control, follow-up answerers
Live answer or AM	0.00422** (0.00209)	0.00300 (0.00263)	0.00861*** (0.00256)	0.00104 (0.00252)	0.00578* (0.00311)	0.00574 (0.00390)	0.0107*** (0.00378)	0.00106 (0.00373)
Constant	0.489*** (0.00136)	0.489*** (0.00136)	0.489*** (0.00136)	0.489*** (0.00136)	0.527*** (0.00257)	0.527*** (0.00257)	0.527*** (0.00257)	0.527*** (0.00257)
Subject controls	No	No	No	No	No	No	No	No
Observations	539,567	271,251	270,157	268,821	150,964	75,904	74,734	75,570
R-squared	5.167e-04	2.517e-04	7.543e-04	8.010e-05	1.539e-04	1.445e-04	2.492e-04	1.365e-05

Note: The table shows LATE estimates for the effect of answering at least one treatment call. Being assigned to the treatment group is used as an instrument for answering a treatment call. The subset of subjects used to estimate each specification is described in the population row. Robust standard errors are displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 16: Differences in voting rates based on the number of treatment live answers

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted
Population	T3, control	T3, control, Follow-up Answerers	T6, control	T6, control, Follow-up Answerers
Treatment and 0 live answers	-0.0469*** (0.00298)	-0.0273*** (0.00946)	-0.0632*** (0.00322)	-0.0829*** (0.0128)
1 live answer	0.0318*** (0.00408)	-0.000320 (0.00777)	0.00269 (0.00536)	-0.0472*** (0.0124)
2 live answers	0.0432*** (0.00392)	0.0180*** (0.00644)	0.0222*** (0.00554)	-0.0177* (0.0108)
3 live answers	0.0511*** (0.00373)	0.0195*** (0.00547)	0.0321*** (0.00531)	0.00455 (0.00901)
4 live answers			0.0427*** (0.00490)	0.0157** (0.00777)
5 live answers			0.0424*** (0.00471)	0.0163** (0.00701)
6 live answers			0.0405*** (0.00447)	0.0128** (0.00619)
Constant	0.489*** (0.00159)	0.527*** (0.00298)	0.489*** (0.00159)	0.527*** (0.00298)
Observations	270,157	74,734	268,821	75,570
R-squared	0.004	0.001	0.004	0.001
F-test	195.9	7.896	119.5	11.44
Prob > F	< 0.001	< 0.001	< 0.001	< 0.001

This table shows the expected difference in voting rate from the control group for subjects in T3 and T6 based on the number of calls that were answered live. Columns 2 and 4 are limited to subjects in households that answered the post-election follow-up call. The subset of subjects used to estimate each specification is described in the population row. The constant represents the voting rate for subjects in the control group. Standard errors are clustered by household and displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 17: County fixed effects intent-to-treat effect among follow-up answerers

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted
Population	Follow-up Answerers	Follow-up Answerers	Follow-up Answerers	Follow-up Answerers
T1		0.0143* (0.00841)		0.0138** (0.00683)
T3		0.0153** (0.00647)		0.0100 (0.00616)
T6		0.0155* (0.00839)		0.0113 (0.00817)
All treatments	0.0148** (0.00658)		0.0123** (0.00577)	
Fixed effects	County FE	County FE	County FE	County FE
Number of fixed effect estimators	562	562	562	562
Subject controls	No	No	Yes	Yes
Observations	146,839	146,839	146,037	146,037
R-squared	9.095e-05	9.128e-05	1.142e-01	1.142e-01
F-test	5.059	2.074	638.7	596.9
Prob > F	0.0249	0.103	< 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 after controlling for county fixed effects. Columns 3 and 4 include subject demographic controls including voter age, gender, education, income, and number of subjects in the household. Data for control variables was not available for a small number of subjects leading to slightly fewer observations than in Table 13. The sample is restricted to follow-up answerers. Standard errors are clustered at the county level and displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 18: City fixed effects intent-to-treat effect among follow-up answerers

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted
Population	Follow-up Answerers	Follow-up Answerers	Follow-up Answerers	Follow-up Answerers
T1		0.0187* (0.00965)		0.0163** (0.00797)
T3		0.0175** (0.00690)		0.0112* (0.00644)
T6		0.0105 (0.00853)		0.00582 (0.00813)
All treatments	0.0166** (0.00770)		0.0125* (0.00674)	
Fixed effects	City FE	City FE	City FE	City FE
Number of fixed effect estimators	4,035	4,035	4,034	4,034
Subject controls	No	No	Yes	Yes
Observations	146,839	146,839	146,037	146,037
R-squared	1.025e-04	1.163e-04	1.172e-01	1.173e-01
F-test	4.637	2.327	568.5	532.0
Prob > F	0.0314	0.0727	< 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 after controlling for city fixed effects. Columns 3 and 4 include subject demographic controls including voter age, gender, education, income, and number of subjects in the household. Data for control variables was not available for a small number of subjects leading to slightly fewer observations than in Table 13. The sample is restricted to follow-up answerers. Standard errors are clustered at the city level and displayed in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table 19: Precinct fixed effects intent-to-treat effect among follow-up answerers

VARIABLES	(1) voted	(2) voted	(3) voted	(4) voted
Population	Follow-up Answerers	Follow-up Answerers	Follow-up Answerers	Follow-up Answerers
T1		0.0357* (0.0204)		0.0215 (0.0202)
T3		0.0443** (0.0216)		0.0297 (0.0210)
T6		0.0621*** (0.0238)		0.0380* (0.0229)
All treatments	0.0443** (0.0179)		0.0276 (0.0177)	
Fixed effects	Precinct FE	Precinct FE	Precinct FE	Precinct FE
Number of fixed effect estimators	17,442	17,442	17,425	17,425
Subject controls	No	No	Yes	Yes
Observations	146,839	146,839	146,037	146,037
R-squared	6.139e-05	7.358e-05	1.259e-01	1.259e-01
F-test	6.151	2.453	779.6	714.7
Prob > F	0.0131	0.0613	< 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 after controlling for precinct fixed effects. Columns 3 and 4 include subject demographic controls including voter age, gender, education, income, and number of subjects in the household. Data for control variables was not available for a small number of subjects leading to slightly fewer observations than in Table 13. The sample is restricted to follow-up answerers. Standard errors are clustered at the precinct level and displayed in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix A: Message Scripts

Calls scripts were used based on the following 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

Script #1:

Hello, this is _____ calling with _____ PAC reminding you to vote on November 4th for our Republican Candidates. Help send a message to Washington that you want your leaders to focus on conservative solutions for Jobs and the Economy. Vote on Tuesday, November 4th for the Republican Ticket.

Thank you. This call was paid for by _____ PAC. Not authorized by any candidate or candidate's committee. ###-###-####

Script #2:

Hello, this is _____ calling with _____ PAC.

Have you, or someone you know, lost your healthcare coverage or seen your premiums skyrocket under Obama's Healthcare Law?

Make your voice heard by voting Republican and letting Congress know you want government out of our healthcare. Remember to vote for our Republican team on Tuesday, November 4th.

Thank you. This call was paid for by _____ PAC. Not authorized by any candidate or candidate's committee. ###-###-####

Script #3:

Hello, this is _____ calling with _____ PAC.

We need your help to stop the gridlock in Washington D.C. and get our government leaders to focus on the issues important to all of us, not just special interest groups.

Send Congress a message that you want conservative leadership who will look out for you. Vote for the Republican Ticket on Tuesday, November 4th.

Thank you. This call was paid for by _____ PAC. Not authorized by any candidate or candidate's committee. ###-###-####

Script #4:

Hello, this is _____ calling with _____ PAC reminding you to vote on November 4th for your Republican Candidates.

With our National Debt spiraling out of control and a weak economy, we need strong candidates more than ever, and this election we have a chance to do something about it.

You can send a message to Washington by voting Republican on Tuesday November 4th.

Thank you. This call was paid for by _____ PAC. Not authorized by any candidate or candidate's committee. ###-###-####

Script #5:

Hello, this is _____ calling with _____ PAC.

This past year, Americans have seen their healthcare premiums skyrocket, while thousands have lost their coverage. We were promised by the President, and nearly every Democrat in Congress, that costs would go down and everyone would be able to keep the coverage they liked. The fact is: they lied.

This Election Day, we can change Washington.

Make your voice heard and tell Congress you want them to get government out of our healthcare by voting for your Republican Ticket on Tuesday, November 4th.

Thank you. This call was paid for by _____ PAC. Not authorized by any candidate or candidate's committee. ###-###-####

Script #6:

Hello, this is _____ calling with _____ PAC.

Today is Election Day and we have a chance to elect Republican candidates who understand that balanced budgets and lower taxes are far better for all of us than increasing the size of government.

Together, we can ensure Congress and our elected leaders focus on what matters.

Be sure to vote for the Republican Ticket today, Tuesday, November 4th.

Thank you. This call was paid for by _____ PAC. Not authorized by any candidate or candidate's committee. ###-###-####

Appendix B: Detailed Treatment Call Outcomes

The following three tables show the number of calls that resulted in a live answer, answering machine, or either one, for subjects (not households) in each treatment group.

Number of calls with live answer	T1		T3		T6		Total	
0	82,565	61%	54,529	40%	43,858	33%	180,952	45%
1	53,355	39%	23,907	18%	12,861	10%	90,123	22%
2	0	0%	26,847	20%	12,079	9%	38,926	10%
3	0	0%	29,543	22%	13,167	10%	42,710	11%
4	0	0%	0	0%	15,382	12%	15,382	4%
5	0	0%	0	0%	16,935	13%	16,935	4%
6	0	0%	0	0%	19,208	14%	19,208	5%
Total	135,920	100%	134,826	100%	133,490	100%	404,236	

Number of calls with AM	T1		T3		T6		Total	
0	89,925	66%	69,137	51%	59,143	44%	218,205	54%
1	45,995	34%	25,626	19%	17,459	13%	89,080	22%
2	0	0%	20,814	15%	14,170	11%	34,984	9%
3	0	0%	19,249	14%	11,528	9%	30,777	8%
4	0	0%	0	0%	10,452	8%	10,452	3%
5	0	0%	0	0%	10,051	8%	10,051	2%
6	0	0%	0	0%	10,687	8%	10,687	3%
Total	135,920	100%	134,826	100%	133,490	100%	404,236	

Number of calls with a live answer or AM	T1		T3		T6		Total	
0	36,570	27%	33,451	25%	31,302	23%	101,323	25%
1	99,350	73%	3,997	3%	2,852	2%	106,199	26%
2	0	0%	4,900	4%	1,886	1%	6,786	2%
3	0	0%	92,478	69%	2,012	2%	94,490	23%
4	0	0%	0	0%	2,890	2%	2,890	1%
5	0	0%	0	0%	4,969	4%	4,969	1%
6	0	0%	0	0%	87,579	66%	87,579	22%
Total	135,920	100%	134,826	100%	133,490	100%	404,236	

The following ten tables show call outcomes for each phone number (household – not subjects) by treatment, call number, and state.

	<u>T1 - Call 1 - Tuesday, Nov. 4th 09:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	5,097	3,441	3,726	4,144	4,037	3,527	23,972	38.41%
AM	3,445	4,556	3,230	2,856	3,556	2,996	20,639	33.07%
Fax	97	138	161	116	104	123	739	1.18%
Busy	97	57	85	66	49	82	436	0.70%
No answer	840	1086	877	836	998	683	5,320	8.52%
Other	834	1,176	2,256	2,311	1,696	3,039	11,312	18.12%
Total	10,410	10,454	10,335	10,329	10,440	10,450	62,418	100.00%

	<u>T3 - Call 1 - Saturday, Nov. 1st 17:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	5,891	4,029	4,377	4,411	4,695	3,834	27,237	43.64%
AM	2,572	3,988	2,453	2,819	2,868	2,202	16,902	27.08%
Fax	98	145	182	110	108	132	775	1.24%
Busy	66	45	66	50	68	109	404	0.65%
No answer	630	1080	958	712	1049	825	5,254	8.42%
Other	1,142	1,170	2,300	2,223	1,653	3,349	11,837	18.97%
Total	10,399	10,457	10,336	10,325	10,441	10,451	62,409	100.00%

	<u>T3 - Call 2 - Monday, Nov. 3rd 18:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	5,002	3,875	3,829	3,889	4,185	3,426	24,206	38.80%
AM	3,507	4,160	3,048	3,378	3,333	2,623	20,049	32.14%
Fax	104	159	203	110	97	133	806	1.29%
Busy	66	62	97	67	80	97	469	0.75%
No answer	561	1013	920	707	960	821	4,982	7.99%
Other	1,156	1,188	2,236	2,171	1,780	3,341	11,872	19.03%
Total	10,396	10,457	10,333	10,322	10,435	10,441	62,384	100.00%

	<u>T3 - Call 3 - Tuesday, Nov. 4th 09:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	5,343	3,490	3,830	4,041	3,844	3,411	23,959	38.42%
AM	3,059	4,473	2,882	3,110	3,491	2,573	19,588	31.41%
Fax	100	151	188	103	112	116	770	1.23%
Busy	74	49	80	58	67	125	453	0.73%
No answer	693	1133	1,098	824	1232	866	5,846	9.37%
Other	1,123	1,157	2,252	2,183	1,685	3,350	11,750	18.84%
Total	10,392	10,453	10,330	10,319	10,431	10,441	62,366	100.00%
	<u>T6 - Call 1 - Thursday, Oct. 30th 18:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	5,627	3,930	4,079	4,169	4,811	3,963	26,579	42.60%
AM	2,688	3,989	2,498	3,073	2,939	2,719	17,906	28.70%
Fax	108	125	147	98	102	129	709	1.14%
Busy	92	64	95	68	52	68	439	0.70%
No answer	679	844	1,085	724	972	567	4,871	7.81%
Other	1,209	1,499	2,429	2,196	1,559	2,995	11,887	19.05%
Total	10,403	10,451	10,333	10,328	10,435	10,441	62,391	100.00%
	<u>T6 - Call 2 - Friday, Oct. 31st 17:30 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	6,028	3,641	4,064	4,223	4,546	3,780	26,282	42.13%
AM	2,255	4,263	2,476	3,019	3,145	2,887	18,045	28.93%
Fax	112	116	149	99	103	123	702	1.13%
Busy	55	59	83	45	59	63	364	0.58%
No answer	732	1143	1,127	689	1010	607	5,308	8.51%
Other	1,220	1,229	2,431	2,252	1,568	2,979	11,679	18.72%
Total	10,402	10,451	10,330	10,327	10,431	10,439	62,380	100.00%

	<u>T6 - Call 3 - Saturday, Nov. 1st 11:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	5,943	3,740	4,158	4,191	4,265	3,518	25,815	41.40%
AM	2,313	4,117	2,364	3,015	3,144	3,102	18,055	28.95%
Fax	102	121	149	94	100	133	699	1.12%
Busy	70	72	99	59	68	62	430	0.69%
No answer	757	714	1,115	716	1034	628	4,964	7.96%
Other	1,210	1,686	2,443	2,246	1,818	2,994	12,397	19.88%
Total	10,395	10,450	10,328	10,321	10,429	10,437	62,360	100.00%

	<u>T6 - Call 4 - Saturday, Nov. 1st 17:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	6,204	3,754	4,331	4,422	4,768	4,015	27,494	44.11%
AM	2,060	4,055	2,172	2,771	2,841	2,611	16,510	26.49%
Fax	95	132	153	100	107	124	711	1.14%
Busy	64	43	118	54	66	96	441	0.71%
No answer	796	1211	1,115	711	1006	622	5,461	8.76%
Other	1,172	1,251	2,437	2,259	1,635	2,965	11,719	18.80%
Total	10,391	10,446	10,326	10,317	10,423	10,433	62,336	100.00%

	<u>T6 - Call 5 - Monday, Nov. 3rd 18:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	4,916	3,763	3,939	3,999	4,273	3,638	24,528	39.38%
AM	3,355	4,126	2,593	3,219	3,303	3,023	19,619	31.50%
Fax	111	129	157	109	95	129	730	1.17%
Busy	117	67	155	78	70	77	564	0.91%
No answer	753	1037	1,094	704	956	575	5,119	8.22%
Other	1,125	1,314	2,382	2,200	1,723	2,984	11,728	18.83%
Total	10,377	10,436	10,320	10,309	10,420	10,426	62,288	100.00%

	<u>T6 - Call 6 - Tuesday, Nov. 4th 09:00 Local</u>							
	VA	PA	GA	OH	NE	NM	Total	Averages
Live	5,511	3,400	3,848	3,952	3,973	3,631	24,315	39.05%
AM	2,599	4,374	2,539	3,125	3,462	2,953	19,052	30.60%
Fax	110	126	152	96	100	132	716	1.15%
Busy	119	75	130	53	72	80	529	0.85%
No answer	916	1207	1,258	846	1172	632	6,031	9.69%
Other	1,120	1,252	2,392	2,228	1,638	2,996	11,626	18.67%
Total	10,375	10,434	10,319	10,300	10,417	10,424	62,269	100.00%

Appendix C: Multiple Hypothesis Testing

Multiple hypothesis testing (MHT) “refers to any instance in which a family of hypotheses is carried out simultaneously and the analyst must decide which hypotheses to reject” (List et al., 2015, page 2). In our analysis, we consider intent-to-treat effects for three distinct treatment groups, T1, T3, and T6. In Section IV. D., we estimated intent-to-treat effects for these three treatment groups among three distinct subgroups – follow-up answerers, subjects who received the follow-up call on an answering machine, and follow-up non-answerers who did not receive the follow-up call on an answering machine. In each of these cases, the fact that we test for an effect for T1, T3, and T6 and do so in the same regression, increases the probability that we would erroneously reject a true null hypothesis. That is, we would erroneously conclude that a treatment was effective even if the apparent effectiveness were only due to random variation in the data.

Following the methodology described in List et al. (2015), we performed MHT adjustments to check the robustness of our findings. Table C1 and Table C2 show MHT-adjusted p-values corresponding to the treatment effects reported in Table 6 and Table 11, respectively.

We find that the statistical significance of the point estimate for T3 that we report in the main text of this paper is robust to MHT adjustments. For example, in our empirical follow-up call analysis, we find that the point estimate for T3 among follow-up answerers remains significant at the five percent level, after MHT adjustments. However, the point estimate for T3 among follow-up non-answerers is no longer statistically significant at the ten percent level after MHT adjustments.

Table C1 – Multiple hypothesis testing for estimates from Table 6

Population/subgroup	Treatment	Difference	Unadjusted p-value	MHT adjusted p-value
All subjects	T1 vs. Control	0.00219	0.2443	0.3937
All subjects	T3 vs. Control	0.00647	0.0003	0.0003
All subjects	T6 vs. Control	0.00080	0.6687	0.6687

Note: This table shows adjusted p-values after correcting for multiple hypothesis testing. Column 1 shows the subject population in the sample, Column 2 shows the intent-to-treat comparison being made, Column 3 shows the difference in voter participation rates between the treatment group and the control group, Column 4 shows the unadjusted estimate of the statistical significance of the estimated treatment effect, and Column 5 shows the adjusted statistical significance of the estimated treatment effect after correcting for MHT.

Table C2 – Multiple hypothesis testing for estimates from Table 13

Population/subgroup	Treatment	Difference	Unadjusted p-value	MHT adjusted p-value
Follow-up answerers	T1 vs. Control	0.00534	0.1453	0.5277
Follow-up answerers	T3 vs. Control	0.01031	0.0033	0.0247
Follow-up answerers	T6 vs. Control	0.00103	0.7867	0.7867
Follow-up AM	T1 vs. Control	0.00318	0.2927	0.7397
Follow-up AM	T3 vs. Control	0.00488	0.1090	0.5133
Follow-up AM	T6 vs. Control	0.00084	0.7823	0.9507
Follow-up Non-answer, no-AM	T1 vs. Control	0.00237	0.4903	0.8737
Follow-up Non-answer, no-AM	T3 vs. Control	0.00738	0.0287	0.1863
Follow-up Non-answer, no-AM	T6 vs. Control	0.00498	0.1357	0.5507

Note: This table shows adjusted p-values after correcting for multiple hypothesis testing. Column 1 shows the subject population in the sample, Column 2 shows the intent-to-treat comparison being made, Column 3 shows the difference in voter participation rates between the treatment group and the control group, Column 4 shows the unadjusted estimate of the statistical significance of the estimated treatment effect, and Column 5 shows the adjusted statistical significance of the estimated treatment effect after correcting for MHT.