Who Is Mobilized to Vote? A Re-Analysis of 11 Field Experiments

Kevin Arceneaux Temple University
David W. Nickerson University of Notre Dame

Many political observers view get-out-the-vote (GOTV) mobilization drives as a way to increase turnout among chronic nonvoters. However, such a strategy assumes that GOTV efforts are effective at increasing turnout in this population, and the extant research offers contradictory evidence regarding the empirical validity of this assumption. We propose a model where only those citizens whose propensity to vote is near the indifference threshold are mobilized to vote and the threshold is determined by the general interest in the election. Our three-parameter model reconciles prior inconsistent empirical results and argues that low-propensity voters can be effectively mobilized only in high-turnout elections. The model is tested on 11 randomized face-to-face voter mobilization field experiments in which we specifically analyze whether subjects' baseline propensity to vote conditions the effectiveness of door-to-door GOTV canvassing. The evidence is consistent with the model and suggests that face-to-face mobilization is better at stimulating turnout among low-propensity voters in prominent elections than it is in quiescent ones.

Political scientists have consistently noted and bemoaned the overall decline in voter turnout that began in the late 1960s and early 1970s (Burnham 1982; Rosenstone and Hansen 1993; Teixeira 1992). Even McDonald and Popkin (2001), who dispute evidence of declining turnout in the 1970s, find a downward trend in participation during the 1960s, which remained lower after 1972 relative to the 1948–70 period. It is intriguing that the secular decline in turnout happened simultaneously with reductions in institutional barriers to vote (e.g., abolition of poll taxes, less restrictive registration laws, motor voter legislation), an increasingly educated electorate, and rising per capita income, each of which is associated with enhancing citizens' propensity to vote (Verba, Schlozman, and Brady 1995; Wolfinger and Rosenstone 1980). Consequently, institutional features and psychological resources, despite their importance in explaining the individual-level likelihood of voting, do not solely account for changes in aggregate participation.

Many scholars attribute the decline in turnout over this period to a shift away from grassroots mobilization by parties and political campaigns (Avery 1989; Rosenstone and Hansen 1993; Teixeira 1987, 1992). To date, an impressive amount of field experimental evidence supports the main assumption underlying this thesis. Building upon early get-out-the-vote (GOTV) field experiments (Eldersveld 1956; Gosnell 1927; Miller, Bositis, and Baer 1981), these studies use random assignment to construct comparable treatment and control groups and produce convincing evidence that door-to-door GOTV drives increase turnout by 7 to 10 percentage points (Gerber and Green 2000; Green, Gerber, and Nickerson 2003). The effectiveness of door-to-door mobilization lies in its use of social psychology to motivate participation. Canvassers are better able to connect with the individuals they visit on a personal level than phone or mail GOTV strategies (Gerber, Green, and Green 2003; Nickerson 2006, 2007). In fact, face-to-face contact is far more effective than more impersonal approaches at motivating a range of behaviors, from blood donation to recycling (Jason et al. 1984; Reams and Ray 1993). Consequently, many citizens decide to vote because someone, in most cases a complete stranger, simply asks them to do so.

Kevin Arceneaux is assistant professor of political science, Temple University, 453 Gladfelter Hall, 1115 West Berks Street, Philadelphia, PA 19122 (kevin.arceneaux@temple.edu). David W. Nickerson is assistant professor of political science, University of Notre Dame, 217 O'Shaughnessy Hall, Notre Dame, IN 46556 (dnickers@nd.edu).

The authors would like to thank James Fowler, Justin Fox, Alan Gerber, Don Green, Alexandra Guisinger, Dan Hungerman, seminar participants at the University of Notre Dame, and anonymous reviewers for helpful comments and suggestions. We would also like to thank Alan Gerber, Don Green, and Melissa Michelson for kindly sharing their data with us. Of course, any errors are our own.

American Journal of Political Science, Vol. 53, No. 1, January 2009, Pp. 1–16

 ISSN 0092-5853

Nevertheless, it is unclear whether GOTV canvassing is effective for everyone, or if some people are easier to mobilize than others. This question has significant political and normative implications, because door-to-door canvassing has been cited by scholars and civic groups as a method to increase participation among perennial nonvoters (Avery 1989). Since the population of nonvoting citizens tends to be poorer and less educated, their abstention is the likely culprit for the lack of correspondence between their opinions and government policy (Bartels 2002). Encouraging participation through GOTV may be a simple and cost-effective way to help alleviate concerns that rising income inequality exacerbates the participation gap between rich and poor, and weaken democratic responsiveness (cf. APSA Task Force Report 2004). Yet if GOTV mobilization is not effective among the population of chronic nonvoters, such optimism may be misplaced.

As we discuss in greater detail in the next section, extant research provides inconsistent evidence regarding the effectiveness of GOTV mobilization among nonvoters. Most of these studies offer less than optimal research designs and measurement, undermining the veracity of their findings. In contrast, our study offers two significant contributions. First, we construct a theory of mobilization that reconciles the prior inconsistent results in a parsimonious three-parameter model. Second, we offer a marked improvement over previous approaches to empirically estimating causal effects by analyzing 11 randomized field experiments. Because GOTV contact is randomized, we are able to circumvent several serious obstacles to causal inference that observational studies cannot. Specifically, we remove the possibility that spurious effects are created by a correlation between one's propensity to vote and strategic campaign targeting, and because we have access to official records of subjects' demographic characteristics and voting behavior in previous elections, we are able to estimate underlying voting propensity with greater precision than current work. We find empirical support for our theoretical model, leading to the conclusion that the effectiveness of GOTV at stimulating participation among chronic nonvoters is contingent on the electoral environment.

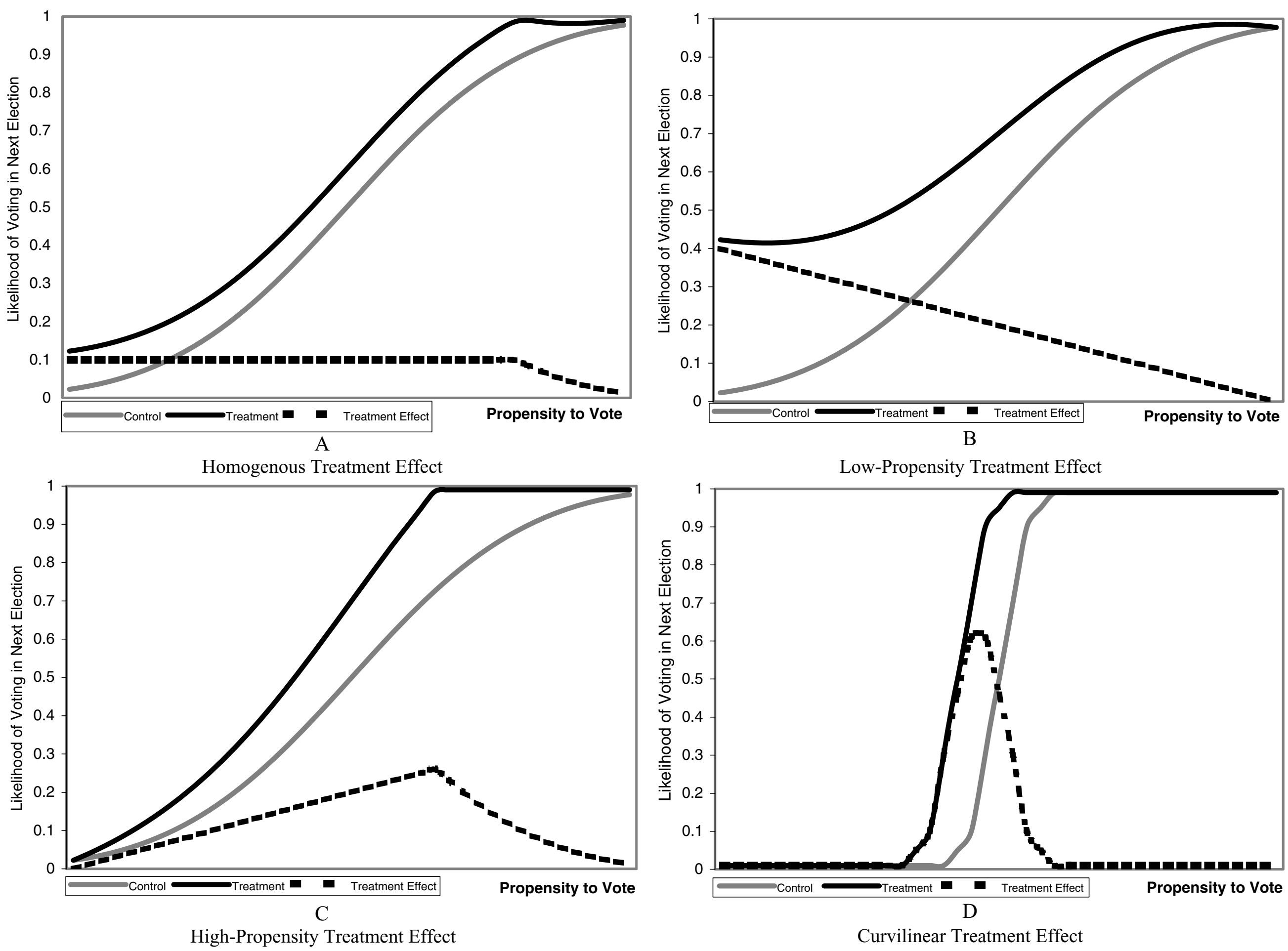
A (Seemingly) Contradictory Literature

There are an infinite number of possible ways in which voting propensity may condition the relationship between GOTV contact and the decision to vote, but the four most prominent possibilities are exposited in Figure 1. Each of the panels in Figure 1 displays three lines. The upper black line is the hypothetical voting rate (y-axis) in the treatment group across levels of voting propensity (x-axis), and the lower gray line is the hypothetical voting rate in the control group. Note that the S-shaped functional form that characterizes the change in voting rates accounts for the fact that the values of the x- and y-axis are bounded by 0 and 1. The black dotted line is the difference between the voting rates between the treatment and control group, also known as the "treatment effect." In sum, GOTV may be equally effective across all types of voters (Panel A), most effective among low-propensity voters (Panel B), among high-propensity voters (Panel C), or among those who fall in the middle of the voting propensity distribution (Panel D).

Extant research finds little evidence for the *homogenous treatment effect* hypothesis, but fails to agree on which of the heterogeneous treatment effects (i.e., Panels B–D) correctly characterizes the relationship between mobilization and voting propensity. Green and Gerber (2004) find that canvassing consistently boosts turnout among individuals who voted in the previous midterm election by twice as much as those who did not. While the experimental design utilized by Green and Gerber ensures campaign outreach is uncorrelated with turnout likelihood, their dichotomous measure of voting propensity is far too coarse to definitively support the *high-propensity treatment effect* model. At the very least, the dichotomous propensity measure only facilitates the estimation of linear relationships making it impossible to detect a *curvilinear treatment effect.*

Hillygus (2005) uses panel survey data to analyze changes in a person's vote intention over the course of the 2000 U.S. presidential campaign, finding some evidence for the *low-propensity treatment effect.* She explains the result by pointing out that individuals who plan on voting cannot be mobilized by campaigns. However, like Green and Gerber (2004), she relies upon a dichotomous measure of voting propensity, thereby excluding a curvilinear relationship. In contrast, Niven (2001, 2004) directly tests and finds support for the *curvilinear treatment effect* hypothesis. Niven argues that politically disengaged individuals will quickly forget campaign messages, while those who regularly vote will not require any persuasion to turn out. He reasons that mobilization can only work on people who lie between the two extremes (2001, 338). Niven provides evidence for his claim by comparing the voter turnout of residents targeted by local campaign to the turnout of individuals not targeted across three different subpopulations: (1) "consistent voters," who voted the past three elections, (2) "intermittent voters," who voted in some but not all of the past three elections, and

FIGURE 1 Hypothesized Relationships between Propensity to Vote and Mobilization



(3) "seldom voters," who did not vote in any of the past three elections. In both studies, Niven finds that door-to-door canvassing is more effective at getting out the vote of those in the intermittent category.

A Theory of Contingent Mobilization

While it is tempting to view these varied findings as competing claims, we offer a single theoretical framework that anticipates all three heterogeneous treatment effects. In short, we argue that the type of voter for whom mobilization is effective is contingent on the electoral context. In a low-salience election, where few people in the electorate are either aware of or interested in the campaign, only high-propensity voters will be receptive to canvassers' blandishments to vote. Conversely, we expect the opposite in a high-salience election. When most people are aware of and interested in an upcoming election, it is mostly those at the low end of the voting propensity spectrum who have not committed to voting. Yet because campaign coverage is intense, even these people have some interest in the election outcome, making them more receptive to entreats to vote than they are in less salient elections. This leaves us with elections of middling interest, such as the typical high-status local election (e.g., mayor) or congressional race. Local news outlets devote some attention to the race, and it is likely that many people are at least aware of the upcoming election and have some interest in the outcome. High-propensity voters are aware and plan to vote, while low-propensity voters are unlikely to be swayed to show up at the polls. As a result, GOTV efforts are more likely to mobilize those who fall in the middle of the voting propensity spectrum in these races.

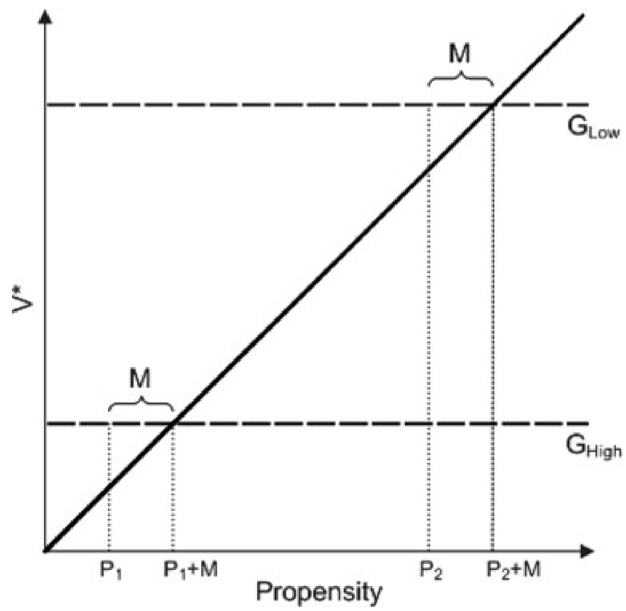
To state our expectations formally, let V_i denote whether an individual votes in the current election, P_i is an individual's latent propensity to vote, M is the effect of any mobilization conducted by the campaign, G is the general interest among the electorate in the election, and V^* is a latent variable that reflects an individual's decision to vote in a particular election. Note that P_i and V^* are distinct but related constructs. One's propensity to vote is an enduring individual-level trait, while the decision to vote is an episodic choice subject to short-term forces. Also note that P_i, M, and G are exogenous to $\mathrm{V_i}$ and V^* (we will say more about this assumption below). We model an individual's decision to vote as a function of his or her underlying propensity and the effect of any GOTV efforts. An individual will vote in the upcoming election, $\mathrm{V_i}$, if V^* surpasses a threshold of interest, which is dictated by G.

$$\mathrm{V}^* = P + M$$
$$\text{and } \mathrm{V} = \begin{cases} 0 & \text{if } V^* \leq -G \\ 1 & \text{if } V^* > -G \end{cases} \tag{1}$$

Figure 2 graphically displays the expectations we derive from equation (1). Voter mobilization should be most clearly observed for those individuals whose propensity to vote places them near the threshold where they are indifferent to voting. The threshold is lower for elections with a great deal of general interest, such as presidential elections, and much higher for elections of little interest, such as school board races. Equation (1) predicts that campaigns should best be able to increase turnout among (a) low-propensity voters during tightly contested, high-profile elections (as depicted in Figure 1B); (b) high-propensity voters during uncompetitive, low-interest elections (Figure 1C); and (c) moderate-propensity voters in elections of middling interest to the general public (Figure 1D).[1]

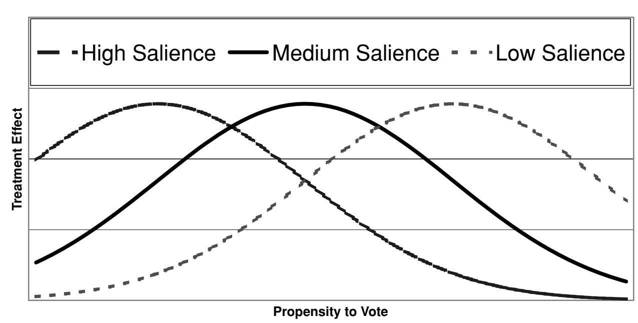
Thus, the theoretical model predicts that the relationship between voting propensity and mobilization should be an inverted-U shape in which the location of the peak of the curve depends on the value of G. In low-salience elections, the peak will be located at the higher end of the voting propensity scale; in medium-salience elections, it will be located near the middle of the propensity scale; and in high-salience elections, it will be located near the lower end of the scale (see Figure 3). As discussed below, the same predictions are derived if we relax the parsimony of the model a bit by adding an error term to account for the probabilistic nature of human decision making (see equations 3 through 5 below).

FIGURE 2 Theoretical Model Predictions



Note: The y-axis, V^*, is the latent likelihood of voting in a particular election. The x-axis, *Propensity*, is a person's enduring baseline propensity for voting. M is the amount voter mobilization increases a person's propensity to vote in an election. G is a cut-point. If a person's propensity to vote is greater than G, then the person votes. The position of G along the y-axis is based upon the salience of the particular election. P_1 and P_2 represent the points at which mobilization can be effective by boosting the propensity to vote over the threshold for voting in two hypothetical elections.

FIGURE 3 Relationship between Propensity to Vote and GOTV Mobilization Predicted by Context-Dependent Theory of Mobilization



[1]Strictly interpreted, equation (1) predicts voter mobilization within a very narrow propensity range. However, there are two primary reasons for expecting a parabolic mobilization effect over a broader range of voters. First and foremost, there will be individual-level idiosyncratic causes of voting that may have the effect of raising or lowering a person's likelihood of voting in the election. The equation and statistical analysis captures average trends rather than individual-level decisions. These individual-level causes will be captured in the estimator's error term. Second, voting propensity will always be measured with uncertainty. The particular point estimate will be accurate within a given range, so mobilization might be observed over a wider range of individuals.

The model supplies an intuitive interpretation of how mobilization works. By personally encouraging people to vote, the campaign hopes to induce poll-avoidant supporters to behave like high-propensity voters. Yet campaigns will not have equal success attracting all low-propensity voters to the polls. Our model anticipates that campaigns successfully achieve this goal only among supporters whose voting propensity is near the threshold set by G. Assuming $G = 0.50$, then, a campaign that can expect to increase, on average, the probability an individual votes by 7 percentage points (i.e., $M = 0.07$) will only have a decisive impact on those who have a 43–50% probability of voting in the election without GOTV contact.[2]

We believe our simple model offers a number of contributions beyond previous research. First, our model of mobilization builds upon considerable empirical evidence that the decision to vote is highly contingent on the electoral context. When the race is close and many people care about the outcome, more people decide to vote relative to races in which general interest is low (e.g., Cox and Munger 1989; Rosenstone and Hansen 1993). Our contribution is to note that the type of voter for whom mobilization is effective must also be contingent on electoral context. As general interest in the campaign increases, mobilization is more likely to reach inveterate nonvoters.

Second, our model is highly consistent with how campaigns actually conceptualize voting behavior when crafting GOTV strategy. Campaigns attempt to mobilize individuals whom they believe will be most receptive to their appeals to vote. They do so by using government voting records to identify individuals who have voted in previous elections. Consequently, campaigns behave as if P_i is exogenous to V^*. Furthermore, a campaign's primary means of boosting turnout is via mobilization, M.[3] Finally, the types of voters targeted by campaigns are affected by G. In high-salience elections, campaigns target unlikely voters out of the belief that everyone else is going to vote without their encouragement, whereas in low-salience elections they assume the opposite and focus on those voters who have reliably voted in the past (Malchow 2003).

In short, our model, as stated in equation (1) and depicted in Figure 2, offers clear predictions about how mobilization interacts with individual voting propensities and electoral context to affect turnout. Not only is our model parsimonious and intuitive, but it also accounts for empirical findings that heretofore have been considered inconsistent and contradictory. Moreover, as we will discuss in the next section, it is possible to measure these parameters in a straightforward and objective fashion, surmounting a major obstacle that has beset attempts to test empirically previous formal models of voter turnout.

Method

The theoretical model presented in equation (1) readily lends itself to an empirical model. By adding G to both sides of the equation and substituting, the mobilization model can be transformed so as to make the cutoff tractable.

$$\mathrm{V}^* = P + M + G$$
$$\text{and } \mathrm{V} = \begin{cases} 0 & \text{if } V^* \leq 0 \\ 1 & \text{if } V^* > 0 \end{cases} \tag{2}$$

When an individual's propensity to vote, P, exposure to mobilization activity, M, and the general salience of the election, G, push the underlying likelihood of voting, V^*, above zero, then the person will vote. Otherwise, the individual will abstain. The mobilization model now parallels commonly used statistical models for dichotomous dependent variables. The parameterization of the models follows in a straightforward manner and is presented in equation (3):

$$\mathrm{V}_\mathrm{i}^* = \beta_0 + \beta_1 P_i + \beta_2 M_i + \beta_3 G_E + \varepsilon_i$$
$$\text{and } \mathrm{V}_\mathrm{i} = \begin{cases} 0 & \text{if } V_i^* \leq 0 \\ 1 & \text{if } V_i^* > 0 \end{cases} \tag{3}$$

where ε_i is a random variable capturing idiosyncratic factors that influence V_i but are unrelated to P_i, M_i, and G_E. All of the quantities of interest in equation (3) vary across individuals and are measured at the individual level (subscripted i) except the salience of the election (subscripted E), which varies only across elections. Thus, standard errors should be much larger for the salience coefficient, β_3, than the other terms.

If one assumes ε_i is distributed normally, equation (3) presents the probit model. Citizen i votes in election

[2] For the sake of modeling parsimony we must assume that we know G with certainty. However, in the empirical estimation of the model, we relax this assumption and include an error term to account for the probabilistic nature of the real world.

[3] Some campaign behaviors, such as issuing press releases, staging events, and conducting muckraking research, can be viewed as affecting G. An analysis of such activities is beyond the scope of this present inquiry.

E if and only if $V_i^* > 0$. Hence, the probability i votes is:

$$\begin{aligned} &\Pr(\beta_0 + \beta_1 P_i + \beta_2 M_i + \beta_3 G_E > -\varepsilon_i) \\ &\quad = \Phi(\beta_0 + \beta_1 P_i + \beta_2 M_i + \beta_3 G_E) \end{aligned} \tag{4}$$

where Φ is the cumulative normal distribution. The effect of being mobilized is then:

$$\begin{aligned} \Delta M_i &= \Phi(\beta_0 + \beta_1 P_i + \beta_3 G_E + \beta_2) \\ &\quad - \Phi(\beta_0 + \beta_1 P_i + \beta_3 G_E) \end{aligned} \tag{5}$$

Thus, ΔM_i represents the effect of being mobilized on person *i*'s probability of voting given *i*'s propensity to vote and the general salience of the election. This function is graphed in Figure 3 and demonstrates the link between the context-dependent theory of mobilization and the estimator presented in equation (3).

The Need for Randomization

Strategic behavior on the part of campaigns and unobserved heterogeneity among individuals could cause M to be correlated with ε, thereby biasing parameter estimates. Campaigns target specific types of individuals, who, in turn, are not equally available to be contacted by campaigns. Political parties also carefully allocate mobilization dollars to particular races based upon complicated decision rules that are a function of G. Furthermore, it is possible that the people targeted and available to receive contact from the campaign possess higher baseline propensities to vote and may be more receptive to blandishments to vote than people who are not available. Thus, strategic behavior on the part of campaigns and individual psychology make observational data unsuitable to test our theory of context-dependent mobilization.[4]

We surmount the potentially problematic behavior of campaigns and individuals by using randomized field experiments. Rather than allowing the campaign to decide whom to target, individual voters are randomly assigned to be contacted, $M = 1$, or not, $M = 0$. Since the assignment is exogenous and random, the treatment group (i.e., people to be contacted) and control group (i.e., people not to be contacted) should possess equal average propensities to vote. That is, $E[Propensity_T - Propensity_C\ ] = 0$ within each experiment.

To have variance in election salience, G, experiments across a range of different elections need to be pooled together. While randomization ensures no correlation between mobilization activities and personal and contextual attributes within experiments, the same cannot be said across experiments. If the proportion of subjects assigned to the control group were systematically smaller in tightly contested elections, then M would be correlated with G and, potentially, ε. Our solution to this problem is to force experiments to be evenly split between treatment and control groups, thereby eliminating any potential biases. For experiments containing more treatment subjects than control, we randomly select treatment subjects to exclude from the experiment to conduct the analysis with an even number of treatment and control subjects from the experiment. The same process is conducted for experiments "overweighted" with control subjects. Eliminating subjects from the analysis is inefficient, but since they are randomly selected it does not introduce bias.[5] In sum, the use of randomized experiments provides unbiased estimates of who is mobilized to vote and corrects a methodological problem in the existing literature.

Average Treatment on the Treated Effect

Like all GOTV efforts, the campaigns in the experiments that we analyze were not able to reach everyone assigned to the treatment group. Some individuals did not answer the door, were not home, or no longer lived at the address listed in the voter files. The failure-to-treat problem does not bias the estimates of the empirical model discussed in the last section, because random assignment ensures that (within sampling variability) the treatment and control group have an equal proportion of contactable

[4]This problem can also be viewed through the lens of measurement error and is illustrated in the studies conducted by Hillygus (2005) and Niven (2001, 2004). Even though Hillygus uses panel data, it is possible that the type of individuals who report being contacted about voting by members in their community are the type of people who tend to vote anyway—*despite the fact that they initially expressed disinterest in voting at the beginning of the campaign.* In Niven's case, an alternative explanation of his findings is that his measure of intermittent voters is biased. Recall that he categorizes an individual as a high-propensity voter if the person voted in *all* of the three past elections. Yet it is likely the case that many individuals who typically vote in elections (i.e., *true* high-propensity voters) could not vote in a previous election because they were ill or away for vacation, causing them to be mislabeled as "intermittent" voters. Meanwhile, it is unlikely that enough true low-propensity voters would accidentally vote in an election so as to counterbalance the miscategorized high-propensity voters. As a result, Niven's measure of intermittent voters will be biased in the direction of finding a turnout boost.

[5]It does, however, introduce additional noise. We take into account this noise by bootstrapping the entire data construction process and reporting the larger bootstrapped standard error.

individuals.[6] However, it does mean that estimates from equation (3) measure the effect of assignment to treatment conditions or the overall effect of GOTV outreach on those that the campaign intended to treat (i.e., the ITT effect), and not the effect on those who were exposed to GOTV contact. The intent-to-treat effect is useful for evaluating the effect of a program (i.e., given that a campaign makes outreach, who responds?), but not estimating the behavioral response of individuals to the actual program intervention (i.e., campaign contact).

An intuitive way to measure the effect of GOTV contact would be to substitute treatment assignment, M, in equation (3) with an indicator for actual GOTV contact, C. Unfortunately, this approach risks introducing bias into the causal estimates since unobservable factors that cause individuals to be exposed to GOTV contact may also be correlated with voting behavior (see Arceneaux, Gerber, and Green 2006 for a demonstration of this point). Instead, it is more appropriate to rework the empirical model in equation (3) by including C as an endogenous function of M. This approach is akin to treating random assignment as an instrument for GOTV contact, which others have shown to be a valid way to estimate average-treatment-on-treated (ATT) effects (Angrist, Imbens, and Rubins 1996; Gerber and Green 2000). Wooldridge (2002, 477–78) offers a blueprint for this approach when both the dependent variable and endogenous explanatory variable are dichotomous, as they are in our case. Using our definitions for variables, the model is:

$$\begin{aligned}
V_i^* &= \beta_0 + \beta_1 P_i + \beta_2 C_i + \beta_3 G_E + \varepsilon_1 \\
V_i &= \begin{cases} 1 & \text{if } V_i^* > 0 \\ 0 & \text{if } V_i^* \le 0 \end{cases} \\
C_i^* &= \gamma_0 + \gamma M_i + \varepsilon_2 \\
C_i &= \begin{cases} 1 & \text{if } C_i^* > 0 \\ 0 & \text{if } C_i^* \le 0 \end{cases}
\end{aligned} \tag{6}$$

where the error terms $(\varepsilon_1, \varepsilon_2) \sim$ bivariate normal and are independent of M. Because M is a random, exogenous variable it is independent of the error terms by design. In order to derive the likelihood function, one must obtain the joint distribution of (V, C) given M. Wooldridge (2002, 478) derives the likelihood function below:

$$\begin{aligned}
\ln\Bigg[& \frac{1}{\Phi(\gamma M)} \int_{-\gamma M}^{\infty} \Phi\big[(\beta_0 + \beta_1 P_i + \beta_2 C_i + \beta_3 G_E \\
& + \omega\varepsilon_2)/(1-\omega^2)^{1/2}\big]\phi(\varepsilon_2)\partial\varepsilon_2 \\
& \times \Bigg(1 - \frac{1}{\Phi(\gamma M)} \int_{-\gamma M}^{\infty} \Phi\big[(\beta_0 + \beta_1 P_i + \beta_2 C_i \\
& + \beta_3 G_E + \omega\varepsilon_2)/(1-\omega^2)^{1/2}\big]\phi(\varepsilon_2)\partial\varepsilon_2\Bigg) \\
& \times \frac{1}{\Phi(\gamma M)} \int_{-\infty}^{-\gamma M} \Phi\big[(\beta_0 + \beta_1 P_i + \beta_2 C_i \\
& + \beta_3 G_E + \omega\varepsilon_2)/(1-\omega^2)^{1/2}\big]\phi(\varepsilon_2)\partial\varepsilon_2 \\
& \times \Bigg(1 - \frac{1}{1-\Phi(\gamma M)} \int_{-\infty}^{-\gamma M} \Phi\big[(\beta_0 + \beta_1 P_i + \beta_2 C_i \\
& + \beta_3 G_E + \omega\varepsilon_2)/(1-\omega^2)^{1/2}\big]\phi(\varepsilon_2)\partial\varepsilon_2\Bigg)\Bigg]
\end{aligned} \tag{7}$$

where $\omega = \mathrm{Corr}(\varepsilon_1, \varepsilon_2)$ and ϕ = probability density function for the normal distribution.

Because V cannot be endogenous to C, this model is a special case of a "recursive, simultaneous-equations" system identified by Greene (2000, 852–53), and the bivariate probit can solve the likelihood equations to estimate the average treatment effect upon those treated (ATT).[7]

Measurement

In addition to avoiding bias from selection processes and unobserved heterogeneity, our analysis seeks to minimize bias stemming from measurement error. To that end, we follow the lead of Gerber and Green (2004) and Niven (2004), who use official voter turnout records to measure the dependent variable, thereby avoiding any bias from self-reported behavior.[8] Similarly, we use campaign records to measure both the assignment to treatment condition and actual contact—again avoiding

[6]The placebo-controlled design avoids this problem altogether, because members in the placebo group were also canvassed, making it possible to compare individuals contacted in the treatment group to individuals contacted in the placebo group.

[7]Bivariate probit likelihood surfaces can be full of saddle points and flat surfaces that make maximization unreliable (Freedman and Sekhon 2008). Our models converged with little difficulty and no signs of problematic local areas. The fact that the treatment-on-the-treated results (i.e., biprobit) mirror the intent-to-treat results (i.e., probit) and that linear two-stage least squares replicates the findings (not reported) gives us added confidence in the bivariate probit analysis.

[8]Hillygus (2005) uses stated vote intention to measure her dependent variable.

self-reporting bias.[9] Thus, our measures of V, M, and C are unproblematic.

Estimating a person's propensity to vote, P, requires more care. Green and Gerber (2004) and Niven (2004) use official voter turnout records to remove measurement error in the dependent variable, but they rely upon turnout in the past one or two elections as a measure of propensity to vote.[10] Consequently, their measures are coarse and unreliable, undermining the ability to make accurate inferences.

Voting propensity can be modeled successfully by taking into account a more complete range of factors. By including common correlates of voting such as age, registration year, party registration, and a more complete voter history, one can derive relatively precise estimates of a person's propensity to vote. As a practical matter, we use factor analysis to collapse all this information into a single dimension measuring a person's propensity to vote, but the analysis is not particularly sensitive to precisely how propensity is modeled.[11]

Two complications arise from this strategy. First, an individual's propensity to vote is obviously an estimate with greater uncertainty surrounding it than our objective measure of campaign contact. To account for this uncertainty, we employ a bootstrap procedure where propensity is reestimated for each sample drawn. Thus, the bootstrapped standard errors account for the variance in our propensity measure. This is an important innovation that improves upon the approach taken in previous research, which is to assume that an individual's voting propensity is measured with certainty.

The second complication is that each propensity estimate is experiment-specific and not comparable across experiments. The particular score generated using factor analysis will depend entirely upon the data used in the estimation, which will vary across electoral settings. In order to make propensities comparable, we standardize each score so that propensity in each experiment has a mean equal to zero and a standard deviation of 1. This approach assumes the subject population in each experiment shares a mean and variance in propensity to vote.[12]

The final term that requires estimation is the general salience of the election, G. Since people must expend some amount of effort to cast a vote, people are more likely to vote as the salience of the election increases (Cox and Munger 1989). Consequently, the level of turnout in an election is an unambiguous expression of the level of citizen interest in the campaign and offers the most direct measure of G. In order to avoid contaminating our measure of G with the effect of the treatment, we use the turnout in the control group (or placebo group) as an indicator of election salience. Because subjects were randomly assigned to the control group, it is appropriate to infer G for the GOTV target universe from turnout in the control group.[13]

Data

Table 1 presents the 11 experiments included in the analysis.[14] Studies were included in the analysis when the data were readily available, the treatment provided was consistent across studies (i.e., face-to-face contact with a very simple "Please vote" message), and the experimental protocol was successfully executed. Data for a range of door-to-door canvassing experiments were collected for two reasons. First, the analysis requires variance in the general salience of the election, G. The experiments included feature turnout ranging from a high of 69% during the extremely close 2000 presidential campaign in Oregon to a low of 8% in an uncompetitive 2001 city council election in Columbus, Ohio (see Table 1, row 6).

Second, while randomized experiments offer unparalleled internal validity with regard to the manipulated

[9]The nature of the panel data forces Hillygus (2005) to rely upon self-reported contact.

[10]Hillygus (2005) relies upon stated vote intention to measure propensity to vote.

[11]Relying only upon past voter history, adding dummy variables to account for neighborhood effects, running a regression to predict turnout in a prior election, principal component, or using the simple percentage of elections voted in produce slightly different coefficients, but substantively the same answer.

[12]An alternative strategy would be to declare particular past elections fixed points by which to construct benchmarks to compare across experiments. One could impose fewer modeling assumptions by not pooling the data at all and measuring treatment effects for different voting propensities within each election. The difficulty in examining the role played by electoral salience, G, is the downside of such a strategy.

[13]Since G is measured at the level of the election rather than the individual and turnout is highly correlated within elections, the number of observations for elections is much closer to the number of elections rather than the number of individuals. To account for this difference, bootstrapping occurs not only among individuals within an experiment but also between elections included in the analysis. The resulting standard errors are much more conservative than standard errors calculated simply using clustering or random effects.

[14]A detailed description of each experiment is available in the original articles.

Table 1 Description of Experiments

City	1 New Haven	2 Eugene	3 Bridgeport	4 Columbus	5 Detroit	6 Minneapolis	7 St. Paul	8 Dos Palos	9 Denver	10 Minneapolis	11 Kansas City
Design	Standard	Standard	Matched	Matched	Matched	Matched	Matched	Matched	Placebo	Placebo	Precinct
Year	1998	2000	2001	2001	2001	2001	2001	2001	2002	2002	2003
Election Day	Yes	By Mail	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Highest Office	Congress	President	School Board	City Council	Mayor	Mayor	Mayor	School Board	Primary	Primary	Ballot Proposition
Overall Turnout (G)	48%	69%	10%	8%	45%	26%	40%	22%	39%	19%	31%
N	31,098	5,062	1,806	2,478	4,954	2,827	2,208	2,186	562	394	9,712
Treatment/Control Ratio	20%	50%	50%	47%	50%	50%	50%	62%	50%	52%	50%
Contact Rate	30%	40%	28%	14%	31%	19%	32%	77%	100%	100%	62%
Intent-to-Treat Effect	2.7% (0.8)	3.4% (1.3)	3.9% (1.5)	1.4% (1.1)	2.6% (1.4)	1.9% (1.6)	4.6% (2.1)	3.1% (1.7)	8.6% (4.2)	10.9% (4.1)	4.4% (2.5)
Average-Treatment-on-Treated Effect	9.8% (3.0)	8.5% (3.3)	13.9% (5.4)	10.0% (8.0)	8.3% (4.6)	10.4% (8.9)	14.4% (6.5)	4.1% (2.2)	8.6% (4.2)	10.9% (4.1)	7.0% (3.9)
Vote History (# Elections)	1	4	2	3	3	14	15	9	7	14	21
Age	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No
Year Registered	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No
Party	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No
Gender	No	No	Yes	No	Yes	No	No	Yes	Yes	Yes	No

Note: Numbers in parentheses report standard errors. Experimental design terminology taken from Nickerson 2005.
Sources: Gerber and Green 2000 (#1); Green and Gerber 2001 (#2); Green, Gerber, and Nickerson 2003 (#3–7); Michelson 2003 (#8); Nickerson 2008 (#9–10); Arceneaux 2005 (#11).

treatment, external validity is nearly always an open question to be answered with future research. Including as many experiments as possible ensures that the average treatment effect reported represents as many communities, elections, and campaigns as possible. The experiments were conducted in large cities like Denver and Minneapolis, medium-sized cities, and a rural area (Dos Palos). The targeted population includes predominantly college students (i.e., Eugene and Columbus), working-class family neighborhoods (e.g., Denver and St. Paul), and low-income neighborhoods (e.g., Bridgeport and Kansas City). While the settings are not fully representative of the United States, they embody characteristics of many communities.

The pooled data set also contains a number of different types of experiments. The protocol design (see Table 1, row 2), size (row 7), contact rate (row 9), and treatment effect (row 10 for ITT and row 11 for ATT) all vary across experiments. As mentioned above, to avoid possible correlation between assignment to the treatment condition and election-level factors, the data need to be winnowed so that the treatment and control groups are equally sized. The only two experiments meaningfully influenced are New Haven and Dos Palos (see row 9). Thus, the results are unlikely to be an artifact of a particular experimental design.

A subject's propensity to vote is measured with varying degrees of precision across the sites. Voter history is abundant in sites like Kansas City and the minimum possible in New Haven (see Table 1, row 12). When available, other variables such as age (row 13), year of registration (row 14), party (row 15), gender (row 16), the number of registered voters in the household, and geography were also used in calculating the propensity to vote. While more data would be preferable, recall that the extant literature on the subject measures propensity using one or three elections. Our measurement, while not perfect, makes use of all the available data and represents a considerable step forward.

TABLE 2 Estimating Theoretical Parameters with Data from 11 Randomized Field Experiments

Variables	ITT Analysis	ATT Analysis
Mobilization (M)	0.12	0.33
	(0.03)	(0.15)
Propensity (P)	0.55	0.54
	(0.07)	(0.07)
Election Salience (G)	3.3	3.3
	(0.3)	(0.3)
Constant	−1.6	−1.7
	(.02)	(.02)
Number of Observations	41,881	41,881
χ^2	11,690.37	17,033.97
Pseudo R^2	0.208	NA

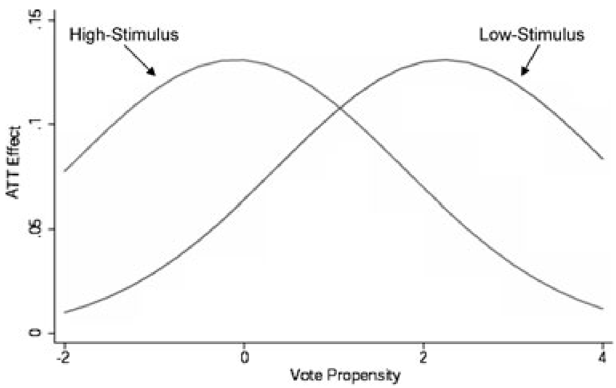
Note: Dependent variable is voter turnout. Coefficients are in probits. Bootstrapped standard errors are in parentheses.

Results

The empirical estimates for the three parameters in our theoretical model are shown in Table 2. Regression results with three variables normally would be highly prone to omitted variable bias, but as we discussed in the last section, because mobilization is randomized in all of these studies, we obviate the need to include control variables. In the first column *M* is estimated with treatment assignment as the measure of attempted mobilization (i.e., the intent-to-treat effect, equation 3). We adjust the estimate of *M* for contact in the second column (i.e., the average-treatment-on-treated effect, equation 7). There are no surprises here. *G* is large, positive, and statistically significant, as it should be. Both *M* and *P* are positive and statistically significant, confirming that door-to-door canvassing increases the marginal probability of voting and that people's underlying propensity to vote is, in fact, actually predictive of whether they do vote. Finally, when we adjust for contact, the size of *M* increases, indicating that individuals who are contacted by door-to-door canvassers are more likely to vote.

Of greater interest to us is what happens to *M* across values of *P* given levels of *G*. We are able to assess this by plugging the estimates displayed in Table 2 into the normal probability distribution and graphing the mobilization effect along the voting propensity continuum for low- and high-salience elections. In order to avoid extrapolating effects that are not supported by our range of data, we use low and high values of *G* that fall within its observed bounds at the 10th and 90th percentile. We display the results of this exercise in Figure 4. As our theory predicts, door-to-door canvassing is most likely to boost turnout among high-propensity voters but not low-propensity voters in low-salience elections ($G = .1$). Conversely, in high-salience elections ($G = .48$), door-to-door canvassing has little effect on high-propensity voters because they are already committed to voting. Instead, it

FIGURE 4 Substantive Effects Generated by Empirical Model



Note: ATT effect estimated from bivariate probit model reported in Table 2, column 2.

is among low-propensity voters that face-to-face contact can be an effective nudge to the polls.

Also, the treatment effect follows a parabolic shape across individuals' propensity to vote, which supports the expectation that mobilization is most effective among those voters who are near the threshold for voting. The location of that threshold depends on the salience of the election. As Figure 4 makes plain, in low-salience elections, door-to-door canvassing has a *de minimus* effect among low-end voters but its effect increases monotonically until it peaks among mid-high-end voters and then begins to taper off. The opposite takes place in high-salience elections. Here, mobilization has a substantial effect even among low-end voters and that effect increases until it peaks among mid-low-end voters. The mobilization effect, then, decreases precipitously and eventually has little effect among high-end voters.

As a robustness check, we excluded studies from the analyses and re-estimated the empirical model. The results we report are not sensitive to the inclusion of any particular study or set of studies. The results also hold up if we step outside of the probit model and include squared terms for P and interaction terms to capture the hypothesized relationships. Because these less parsimonious models with no theoretical motivation yield similar results, we conclude the probit model is not imposing the relationship on the data (see the appendix for tables and details).

Two implications of our findings deserve comment. First, as the figures make clear, treatment-on-treated effects range from 1.3 to 13.2 percentage points. Thus, the 7 to 10 percentage point average treatment effect from face-to-face contact reported by previous field experiments (cf. Green, Gerber, and Nickerson 2003) conceals a great deal of heterogeneity in response to campaign contact. Second, because these effects vary a great deal across individuals and electoral contexts, the cost per vote that a campaign can expect depends on whom they try to mobilize and in what election the mobilization takes place.

To make this last point concrete, we present a cost-per-vote analysis in Table 3. The ingredients for this analysis are straightforward:

$$\$\ per\ vote = \frac{\$HR}{(ATT * \#DK * CR)} \tag{8}$$

where $\$HR$ = hourly rate paid to canvassers, $\#DK$ = number of door knocks that a canvasser can expect to complete in an hour, and CR = the contact rate. Obviously, these quantities will vary depending on the specific situation, but we can nevertheless illustrate the implications of our empirical model by referring to the available data and inserting plausible numbers. We estimate the contact rate for low-, medium-, and high-propensity voters by regressing contact on vote propensity and calculating the probability that a voter will be contacted at a given propensity.[15] These estimates are shown in column 1 of Table 3. As one would expect, the probability of contact is positively correlated with voting propensity. The ATT effect for each of these voters is displayed in columns 2 and 3 of Table 3 for both low- and high-salience elections. These estimates come from the bivariate probit model and are merely a tabular display of the results shown in Figure 4.[16] Finally, we assume that canvassers are paid \$15 an hour and can knock on 25 doors per hour.[17]

The cost per vote is shown in columns 4 and 5 of Table 3 and the results are striking. In elections that generate little interest, a campaign fitting our description would spend approximately \$93 per vote on average if it targets low-propensity voters, but only a modest \$16 per vote if it focuses on high-propensity voters. These results reverse themselves in an election that generates a high degree of interest. For high-turnout elections, high-propensity voters cost over \$60 a vote and low-propensity voters are a far more reasonable \$25 per vote. Of course, these cost estimates change when the particular campaign context differs from our assumptions. For instance, a nonprofit

[15]We employed probit to estimate this model and set $P = -1$ for the low-propensity voter, 1 for the medium-propensity voter, and 3 for the high-propensity voter.

[16]We calculated these figures by estimating the treatment effect in the bivariate probit model for low, medium, and high values of P (see footnote 15) with G set to 0.1 for the low-salience condition and 0.48 for the high-salience condition.

[17]These assumptions are very similar to those made by Green and Gerber (2004) in their cost-per-vote analysis.

TABLE 3 Cost-per-Vote Analysis

		Treatment Effect on the Treated		Dollars per Vote	
Propensity	Odds of Contact	Low-Salience Election	High-Salience Election	Low-Salience Election	High-Salience Election
Low	19.5%	3.3%	12.4%	$93.24	$24.81
Medium	24.3%	11.7%	11.6%	$21.10	$21.29
High	29.7%	12.3%	3.2%	$16.42	$63.13

Note: Dollars per Vote $= \frac{\text{Wage}}{\text{Doors per hour} * \text{Contact Rate} * \text{Treatment Effect}}$.
Wage rate assumed to be $15 per hour, and workers are assumed to knock on 25 doors per hour.

organization that can rely on volunteer labor may be able to drastically reduce its expected cost per vote.[18] Similarly, an organization that can obtain a higher contact rate would also be able to mobilize voters more cheaply. Nevertheless, our analysis strongly demonstrates that campaigns should not adopt a one-size-fits-all mentality for their GOTV operation. As we will touch on in more detail below, the strategy a campaign chooses is contingent on the interaction between its goals and the election context it confronts.

Conclusion

These experimental data offer evidence in support of our contextual model of voter mobilization. Door-to-door canvassing increases turnout mostly by enticing those who are on the cusp of voting. If asked, these fence-sitters are more likely to vote. However, the location of the fence changes depending on the electoral context. In high-profile, competitive races, most registered voters will be above the threshold for voting and mobilization will not be cost-effective. In low-salience, uncompetitive elections, face-to-face conversations about the importance of voting will not be sufficient to bring unlikely voters to the polls. In these cases, a campaign's efforts are best directed towards high-propensity voters, who might be persuaded that voting in "minor" elections is a worthwhile endeavor.

These findings offer a little encouragement for those who want to engage traditionally low-turnout groups in politics. A cost-effective means of increasing voter participation among these groups in sleepy local elections is unlikely to present itself. However, such groups can be mobilized in high-profile elections, such as presidential elections. Since voting may be habitual (Gerber, Green, and Shachar 2003), perhaps a culture of voting can be generated and low-propensity voters can gradually be turned into moderate-propensity voters. Our results suggest that the best place to begin the transformation process is high-salience elections (e.g., Republicans targeting 4 million low-turnout Evangelical voters during the 2004 presidential election).

Our results also offer support for models that characterize the voting decision as an exercise in weighing costs and benefits (e.g., Riker and Ordeshook 1968). Some scholars have argued that because these models do not adequately predict the level of turnout, they are of little analytical use (Green and Shapiro 1994). Yet, these data solidly support one prediction made by this general approach: it should be easier to mobilize those for whom the costs of voting slightly outweigh the benefits of voting relative to those for whom the benefits are significantly outweighed by the costs.

Nevertheless, our theory of contingent mobilization offers a more useful conceptualization of how mobilization influences turnout than the classic Riker and Ordeshook (1968) framework, which has often been invoked when theorizing about the effects of campaigns on voter turnout (e.g., Rosenstone and Hansen 1993). While our model is consistent with the Riker and Ordeshook model, it replaces abstract terms, such as duty and the costs of voting, with more concrete concepts that can be measured. A major weakness of the Riker and Ordeshook formulation is the practical inability to disentangle its various components when contemplating how campaign mobilization boosts turnout. A GOTV campaign could be effective via the benefit term by persuading individuals to care about the outcome of the election, through the cost term by lowering the informational costs associated with voting, through the duty term by reminding people about their role as democratic citizens, or even through the infamous p-term by erroneously convincing individuals that their vote could be decisive.

[18]Its ability to cut labor costs depends on how much money is required to recruit a stable pool of volunteers.

Furthermore, it is unclear how the Riker and Ordeshook model explains why voting propensity or the electoral context conditions voting behavior. Do high-propensity voters have a large sense of duty or low costs associated with voting? Do presidential elections, in which there is typically a great deal of general interest, raise turnout by lowering information costs and boosting notions of civic duty or by increasing the perceived benefits of electing the preferred candidate? Most relevant aspects of elections and political campaigns contribute to both the costs and the psychological benefits of voting, making the calculus of voting of little practical value in guiding campaigns or predicting individual responses to external stimuli. In contrast, we believe that our model offers a parsimonious, intuitive, and testable explanation of voter mobilization. The propensity to vote (P), salience of the election (G), and effect of contact from the campaign (M) can all be measured, and the model accurately predicts our experimental findings.

While our model captures the dynamics of mobilization from a campaign's perspective, future work should focus on the microfoundations underlying the heterogeneity in voting propensities that have been taken as exogenous, both here and in the literature. In particular, our model is entirely static and the propensity to vote, P, is viewed as a fixed attribute. As mentioned above, there is solid evidence that voting is habit forming (Gerber, Green, and Shachar 2003), so propensity in one election is a function of mobilization in a past election. That is, $P_{i,t} = f(P_{i,t-1}, M_{i,t-1})$. In this manner, the dynamics of mass voter behavior might be modeled appropriately. Our parsimonious model of mobilization offers a useful framework upon which future endeavors can build.

There is no reason to believe that the model's utility is restricted to voter turnout. A broad swath of civic engagement can be explained by the context-dependent model. For instance, an individual probably has an underlying propensity to donate to a candidate (P). Requests from campaigns to donate money to the candidate make the person more likely to donate (M), but may not push him or her over the threshold to donate. This threshold (G) depends upon the profile of the cause, election, or candidate. Candidates in close, high-profile elections probably have a lower threshold for donations. In this way, we think the model can be easily adapted to other forms of behavior such as campaign donations, letter writing, volunteer work, protests, parental involvement in schools, and any other activity where an individual bears a cost for a collective good.

Finally, it is worth noting that these experiments focus on individuals who are already registered to vote. The evidence here has little to say about the effectiveness of first registering and then mobilizing individuals who are not registered to vote. GOTV organizations typically focus mobilization efforts on registered voters since official voting records from which target walk lists are constructed are easy to obtain. Moreover, there have been few, if any, field experiments conducted to date on the effectiveness of registering citizens to vote. Consequently, little is known about the difficulties and costs associated with identifying and contacting unregistered citizens. As scholars continue studying ways in which to boost turnout among nonvoters, we believe that the mobilization of unregistered citizens is a significant missing puzzle piece and warrants scrutiny.

Appendix

In order to be sure that our results are not driven by any particular study or group of studies, we re-estimated the probit model after successively removing each study, or excluding groups of studies, such as the placebo-controlled, low-contact, and high-contact experiments. These estimates, reported in Table A1, show that our findings are robust across sample restrictions.

We also want to be sure that the probit model, which perfectly captures our theoretical model, is not driving the results by imposing the hypothesized functional form on the data. We do so by adopting the linear probability model and testing for the parabolic function by including a squared term for P and interacting it with M. If, on average, the treatment effect increases across values of P until reaching a tipping point, the coefficient for $M * P^2$ should be negative. Using two-stage least squares regression with random assignment as an instrument for campaign contact (cf. Angrist, Imbens, and Rubin 1996; Gerber and Green 2000), we find this to be the case for the full sample and for a host of sample restrictions. These results are shown in Table A2, column 1.

Finally, we want to be sure that the apex of the parabola that represents the treatment effect across voting propensity moves in response to changes in G as hypothesized. We can test this in the linear probability framework by interacting G with all the variables in the model reported in Table A2. This approach is equivalent to a hierarchal linear model with G conditioning the effect that M and P have on turnout (cf. Steenbergen and Jones 2002). If the apex of the parabola moves as we expect, the coefficient for $M * P * G$ should be negative. The results in Table A2, column 2 show that this is the case across all of the sample restrictions. We graph the results for the full sample model in Figure A1 to demonstrate findings are commensurate with those obtained from the probit models.

Table A1 Robustness Checks for Probit Model

	Results After Removing:										
Variables	Bridgeport	Columbus	Denver	Dos Palos	Eugene	Kansas City	Minneapolis	Minneapolis Placebo	New Haven	St. Paul	Placebo Experiments
M	0.125	0.127	0.125	0.125	0.129	0.118	0.131	0.125	0.135	0.127	0.123
	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.016)	(0.014)	(0.014)	(0.016)	(0.014)	(0.014)
P	0.565	0.552	0.553	0.561	0.540	0.526	0.539	0.555	0.627	0.542	0.551
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)
G	3.209	3.088	3.243	3.249	3.330	3.326	3.240	3.250	3.372	3.233	3.245
	(0.049)	(0.049)	(0.046)	(0.047)	(0.062)	(0.050)	(0.047)	(0.046)	(0.050)	(0.046)	(0.046)
Constant	−1.602	−1.547	−1.620	−1.621	−1.646	−1.671	−1.621	−1.622	−1.666	−1.618	−1.620
	(0.022)	(0.023)	(0.021)	(0.022)	(0.025)	(0.024)	(0.022)	(0.021)	(0.022)	(0.021)	(0.021)
Observations	55,982	55,308	57,224	55,188	52,729	48,074	54,960	57,392	32,586	55,585	56,830
χ^2	12,445.87	11,621.92	13,137.22	12,627.00	10,606.16	10,722.08	12,210.26	13,140.90	10,652.10	12,666.90	12,963.66
Pseudo R^2	0.16	0.15	0.17	0.17	0.15	0.16	0.16	0.17	0.25	0.17	0.17

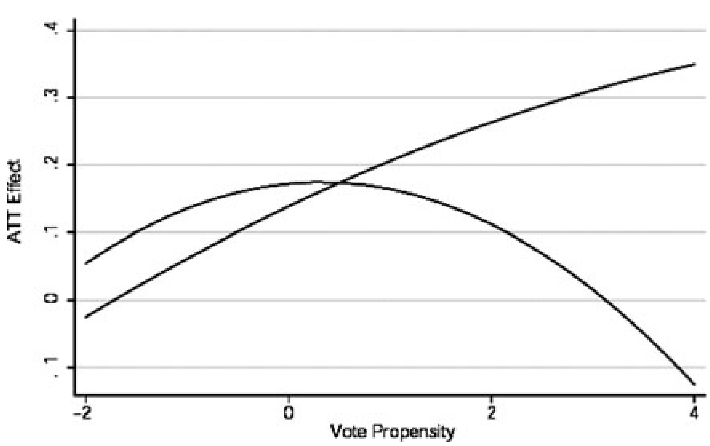
	Results After Excluding:	
Variables	High Contact Only	Low Contact Only
M	0.108	0.177
	(0.016)	(0.026)
P	0.520	0.653
	(0.008)	(0.013)
G	3.357	4.013
	(0.053)	(0.332)
Constant	−1.686	−1.813
	(0.026)	(0.096)
Observations	29,220	12,643
χ^2	8,324.757	2,885.253
Pseudo R^2	0.208	0.186

TABLE A2 Robustness Checks on Model Specification

	Interacted with Propensity	Fully Interacted
M	0.112	0.079
	(0.014)	(0.038)
P	0.192	0.130
	(0.004)	(0.010)
P^2	−0.019	0.050
	(0.003)	(0.007)
G		1.210
		(0.025)
$M * P$	0.028	0.087
	(0.015)	(0.036)
$M * P^2$	−0.015	−0.0003
	(0.010)	(0.026)
$M * G$		0.088
		(0.096)
$P * G$		0.165
		(0.022)
$P^2 * G$		−0.207
		(0.017)
$M * P * G$		−0.152
		(0.086)
$M * P^2 * G$		−0.046
		(0.068)
Constant	0.390	−0.056
	(0.004)	(0.010)
Observations	41,863	41,863
F	1,417.512	1,350.546
Adjusted R^2	0.150	0.260

Model: Linear two-stage least squares.
Dependent variable: Dichotomous voter turnout.
Standard errors reported in parentheses.

FIGURE A1 Linear Probability Model Substantive Effects



References

American Political Science Association Task Force. 2004. "American Democracy in an Age of Rising Inequality." *Perspectives on Politics* 2(December): 651–66.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(June): 444–55.

Arceneaux, Kevin. 2005. "Using Cluster Randomized Field Experiments to Study Voting Behavior." *Annals of the American Academy of Political and Social Science* 601(September): 169–79.

Avery, Michael. 1989. *The Demobilization of American Voters*. New York: Greenwood.

Bartels, Larry M. 2002. "Economic Inequality and Political Representation." Unpublished manuscript. Woodrow Wilson School of Public and International Affairs, Princeton University.

Burnham, Walter Dean. 1982. *The Current Crisis in American Politics*. New York: Oxford University Press.

Cox, Gary W., and Michael C. Munger. 1989. "Closeness, Expenditures, and Turnout in the 1982 U.S. House Elections." *American Political Science Review* 83(March): 217–31.

Eldersveld, Samuel J. 1956. "Experimental Propaganda Techniques and Voting Behavior." *American Political Science Review* 50(March): 154–65.

Freedman, David A., and Jasjeet S. Sekhon. 2008. "Endogeneity in Probit Response Models." Unpublished manuscript. Downloaded from http://www.stat.berkeley.edu/∼census/heckprob.pdf on July 31, 2008.

Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Personal Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94(September): 653–64.

Gerber, Alan S., Donald P. Green, and Matthew N. Green. 2003. "The Effects of Partisan Direct Mail on Voter Turnout." *Electoral Studies* 22(December): 563–79.

Gerber, Alan S., Donald P. Green, and Ron Shachar. 2003. "Voting May Be Habit Forming: Evidence from a Randomized Field Experiment." *American Journal of Political Science* 47(July): 540–50.

Green, Donald P., and Alan S. Gerber. 2001. "Getting Out the Youth Vote: Results from Randomized Field Experiments." Unpublished manuscript. Yale University. http://www.yale.edu/isps/publications/youthvole.pdf.

Green, Donald P., and Alan S. Gerber. 2004. *Get Out the Vote! How to Increase Voter Turnout*. Washington, DC: Brookings Institution Press.

Green, Donald P., Alan S. Gerber, and David W. Nickerson. 2003. "Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments." *Journal of Politics* 65(November): 1083–96.

Green, Donald P., and Ian Shapiro. 1994. *The Pathologies of Rational Choice*. New Haven, CT: Yale University Press.

Greene, William H. 2000. *Econometric Analysis*. 4th ed. Upper Saddle River, NJ: Prentice-Hall.

Gosnell, Harold F. 1927. *Getting-Out-the-Vote: An Experiment in the Stimulation of Voting*. Chicago: University of Chicago Press.

Hillygus, D. Sunshine. 2005. "Campaign Effects and the Dynamics of Turnout Intention in Election 2000." *Journal of Politics* 67(February): 50–68.

Jason, Leonard A., Tomas Rose, Joseph R. Ferrari, and Russ Barone. 1984. "Personal versus Impersonal Methods for Recruiting Blood Donations." *Journal of Social Psychology* 123(June): 139–40.

Malchow, Hal. 2003. *The New Political Targeting*. Washington, DC: Campaigns and Elections.

McDonald, Michael P., and Samuel L. Popkin. 2001. "The Myth of the Vanishing Voter." *American Political Science Review* 95(December): 963–74.

Michelson, Melissa R. 2003. "Getting Out the Latino Vote: How Door-to-Door Canvassing Influences Voter Turnout in Rural Central California." *Political Behavior* 25(September): 247–63.

Miller, Roy E., David A. Bositis, and Denise L. Baer. 1981. "Stimulating Voter Turnout in a Primary: Field Experiment with a Precinct Committeeman." *International Political Science Review* 2(October): 445–60.

Nickerson, David W. 2005. "Scalable Protocols Offer Efficient Design for Field Experiments." *Political Analysis* 13(Summer): 233–52.

Nickerson, David W. 2006. "Volunteer Phone Calls Can Increase Turnout: Evidence from Eight Field Experiments." *American Politics Research* 34(May): 271–92.

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

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

Niven, David. 2001. "The Limits of Mobilization: Turnout Evidence from State House Primaries." *Political Behavior* 23(December): 335–50.

Niven, David. 2004. "The Mobilization Solution? Face-to-Face Contact and Voter Turnout in Municipal Elections." *Journal of Politics* 66(August): 868–85.

Reams, Margaret A., and Brooks H. Ray. 1993. "The Effects of Three Prompting Methods on Recycling Participation Rates: A Field Study." *Journal of Environmental Systems* 22(4): 371–79.

Riker, William H., and Peter C. Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Political Science Review* 62(March): 25–42.

Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, Participation, and Democracy in America*. New York: Macmillan.

Schenker, Nathaniel, and Jane F. Gentleman. 2001. "On Judging the Significance of Differences by Examining Overlap between Confidence Intervals." *The American Statistician* 55(August): 182–86.

Teixeira, Ruy. 1987. *Why Americans Don't Vote*. New York: Greenwood.

Teixeira, Ruy. 1992. *The Disappearing American Voter*. Washington, DC: Brookings.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* New Haven, CT: Yale University Press.

Wooldridge, Jeffry M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.