

**Who is Mobilized to Vote?
A Re-Analysis of Eleven Randomized Field Experiments**

Kevin Arceneaux
Assistant Professor
Department of Political Science
Temple University
453 Gladfelter Hall
1115 West Berks Street
Philadelphia, PA 19122
kevin.arceneaux@temple.edu

David W. Nickerson
Assistant Professor
Department 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, Alan Gerber, Don Green, Alexandra Guisinger, Sunshine Hillygus, 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.

Abstract

Many political observers view get-out-the-vote (GOTV) mobilization drives as a way to increase turnout among chronic non-voters. 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 1948-1970 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 are associated with enhancing citizens' propensity to vote (Verba, Scholzman, 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 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 get-out-the-vote (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 (Nickerson 2006; Gerber, Green and Green 2003). 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.

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 non-voters (Avery 1989). Since the population of non-voting 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 non-voters, 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 non-voters. 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 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 non-voters 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 explicated 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).

[Figure 1 about here]

Extant research finds little evidence for the *homogenous treatment effect* hypothesis, but fails to agree on which of the heterogeneous treatment effects (i.e., Panel 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 mid-term 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 US 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, 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 turnout. 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 sub-populations: 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 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 underlying propensity to vote, M is the effect of any mobilization conducted by the campaign, and 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. Note that P_i , M , and G are exogenous to V_i and V^* (we will say more about this assumption below). We model an individual's decision to vote as a function of their underlying propensity and the effect of any

GOTV efforts. An individual will vote in the upcoming election, V_i , if V^* surpasses a threshold of interest, which is dictated by G .

$$V^* = P + M \tag{1}$$

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

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

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 wider range of individuals.

scale (see Figure 3). As we discuss below, we also derive the same predictions 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).

[Insert Figure 3 about Here]

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 percent (i.e., $M = 0.07$), will only have a decisive impact on those for whom have a 43 to 50 percent 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 non-voters.

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.

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).

Third, our model is 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.

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.

Furthermore, it is unclear how their 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, 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. Our not only is our model parsimonious and intuitive, 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 cut-off tractable.

$$V^* = P + M + G \tag{2}$$

$$\text{and } V = \begin{cases} 0 & \text{if } V^* \leq 0 \\ 1 & \text{if } V^* > 0 \end{cases}$$

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:

$$V_i^* = \beta_0 + \beta_1 P_i + \beta_2 M_i + \beta_3 G_E + \varepsilon_i \quad (3)$$

$$\text{and } V_i = \begin{cases} 0 & \text{if } V_i^* \leq 0 \\ 1 & \text{if } V_i^* > 0 \end{cases}$$

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

$$\Pr(\beta_0 + \beta_1 P_i + \beta_2 M_i + \beta_3 G_E > -\varepsilon_i) = \Phi(\beta_0 + \beta_1 P_i + \beta_2 M_i + \beta_3 G_E) \quad (4)$$

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

$$\Delta M_i = \Phi(\beta_0 + \beta_1 P_i + \beta_3 G_E + \beta_2) - \Phi(\beta_0 + \beta_1 P_i + \beta_3 G_E) \quad (5)$$

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 makes 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

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 individual who reports being contacted about voting by members in their community are the type of people who tend to vote anyway – *despite the fact that he 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.

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 “over-weighted” 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 individuals.⁶ However, it does mean that estimates from equation 3 measure the effect of assignment to treatment conditions or the overall effect of

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 .

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.

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:

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

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:

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

where $\omega = \text{Corr}(\varepsilon_1, \varepsilon_2)$ 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 it means we can use bivariate probit to solve the likelihood equations and estimate the average treatment effect upon those treated (ATT).

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.⁷ Similarly, we use campaign records to measure both the assignment to treatment condition and actual contact – again avoiding self-reporting bias.⁸ Thus, our measures of V , M , and C are unproblematic.

⁷ Hillygus (2005) uses stated vote intention to measure her dependent variable.

⁸ Once again, the nature of the panel data forces Hillygus (2005) to rely upon self-reported contact.

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.⁹ The minimal number of elections used means the measure of propensity is coarse, unreliable, and make inference difficult.

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, neighborhood, 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.¹⁰

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 re-estimated 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

⁹ Hillygus (2005) relies upon stated vote intention to measure propensity to vote.

¹⁰ Relying only upon past voter history, running a regression to predict turnout in a prior election, or using the simple percentage of elections voted in produce slightly different coefficients, but substantively the same answer.

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 1.¹¹

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.¹²

Data

Table 1 presents the eleven experiments included in the analysis. (Reviewer's Appendix A contains a detailed description of each experiment.) Data for every available door-to-door canvassing experiment was 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 percent during the extremely close 2000 Presidential campaign in Oregon to a low of 8 percent in an uncompetitive 2001 city council election in Columbus, OH (see Table 1, row 6).

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

$$^{11} \hat{P}_{std} = \frac{\hat{P} - \bar{P}}{sd_{\hat{P}}}$$

¹² Since G is measured at the level of the election rather than the individual, all analyses will need to cluster standard errors on the election to account for correlation within elections and the smaller functional N .

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 are in many ways typical of most communities.

The pooled data set also contain a number of different types of experiments. The protocol design (see Table 1, row 3), size (row 8), 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.

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.

[Table 2 about here]

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

[Figure 4 about here]

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. The shaded areas around each line depict the 95-percent confidence interval calculated from the bootstrap simulations. They demonstrate graphically that the treatment effects are statistically distinguishable from zero, and that the low- and high-salience lines are statistically different from one another, especially at the maxima and tails.¹³

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 Reviewers' Appendix B for tables and details.)

¹³ Also, note that when comparing two regression lines using a confidence interval overlap test, the 95-percent confidence interval produces a conservative test (i.e., $\alpha = 0.006$, not 0.05) (Schenker and Gentleman 2001). If we use the appropriate confidence interval for the standard 0.05 Type I error rate, the difference between the two lines is even more apparent. These results are available from the authors upon request.

Two implications of our findings deserve comment. First, as the figures make clear, treatment-on-treated effects range from 1.3 percent to 13.2 percent. 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:

$$\text{\$ per vote} = \frac{\text{\$}HR}{(ATT * \#DK * CR)} \quad (9)$$

where $\text{\$}HR$ = 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.¹⁴ 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

¹⁴ 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.

shown in Figure 4.¹⁵ Finally, we assume that canvassers are paid \$15 an hour and can knock on 25 doors per hour.¹⁶

[Table 3 about here]

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 organization that can rely on volunteer labor may be able to drastically reduce their expected cost per vote.¹⁷ 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 impressive evidence in support of our contextual model of voter mobilization. Door-to-door canvassing increases turnout mostly by enticing those who are

¹⁵ We calculated these figures by estimating the treatment effect in the bivariate probit model for low, medium, and high values of P (see endnote 13) with G set to 0.1 for the low-salience condition and 0.48 for the high-salience condition.

¹⁶ These assumptions are very similar to those made by Green and Gerber (2004) in their cost-per-vote analysis.

¹⁷ Their ability to cut labor costs depends on how much money is required to recruit a stable pool of volunteers.

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.

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. Our parsimonious model of mobilization offers a useful framework upon which future endeavors can build.

Also, there is no reason to believe that the model's utility is restricted to voter turnout. We believe that 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 makes the person more likely to donate (M), but may not push them 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.

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.
- 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): 540-550.
- 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: Yale University Press.
- Greene, William H. 2000. *Econometric Analysis, 4th Edition*. Upper Saddle River, New Jersey: 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. 2006. "Volunteer Phone Calls Can Increase Turnout: Evidence from Eight Field Experiments." *American Politics Research*, 34 (May): 271-92.
- Nickerson, David W. 2004. "Is Voting Contagious?" Unpublished manuscript. Yale University.
- Nickerson, David W. 2005. "Scalable Protocols Offer Efficient Design for Field Experiments." *Political Analysis*, 13 (Summer): 233-252.
- Niven, David. 2004. "The Mobilization Solution? Face-to-Face Contact and Voter Turnout in Municipal Elections." *Journal of Politics*, 66 (August): 868-85.
- Niven, David. 2001. "The Limits of Mobilization: Turnout Evidence from State House Primaries." *Political Behavior*, 23 (December): 335-50.
- Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- 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): 371-9.

- 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. 1992. *The Disappearing American Voter*. Washington, DC: Brookings.
- Teixeira, Ruy. 1987. *Why Americans Don't Vote*. New York: Greenwood.
- Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.
- Wooldridge, Jeffry M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* New Haven, CT: Yale University Press

Table 1: Description of Experiments

	1	2	3	4	5	6	7	8	9	10	11
City	New Haven	Eugene	Bridgeport	Columbus	Detroit	Minneapolis	St. Paul	Dos Palos	Denver	Minneapolis	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	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.

Sources: Gerber and Green 2000 (#1); Green, Gerber and Nickerson 2003 (#3-7); Michelson 2003 (#8); Nickerson 2005 (#9); Arceneaux 2005 (#11)

Table 2: Estimating Theoretical Parameters with Data from 11 Randomized Field Experiments

Variables	ITT Analysis	ATT Analysis
Mobilization (M)	0.127 (0.013)	0.352 (0.033)
Propensity (P)	0.557 (0.008)	0.554 (0.008)
Election Salience (G)	3.247 (0.169)	3.305 (0.144)
Constant	-1.622 (.020)	-1.658 (.021)
Number of Observations	41,863	41,863
χ^2	11,704.37	16,879.26
Pseudo R ²	0.209	NA

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

Table 3: Cost-Per-Vote Analysis

Propensity	Odds of Contact	Treatment Effect on the Treated		Dollars per Vote	
		Low-Saliency Election	High-Saliency Election	Low-Saliency Election	High-Saliency 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.

Figure 1: Hypothesized Relationships between Propensity to Vote and Mobilization

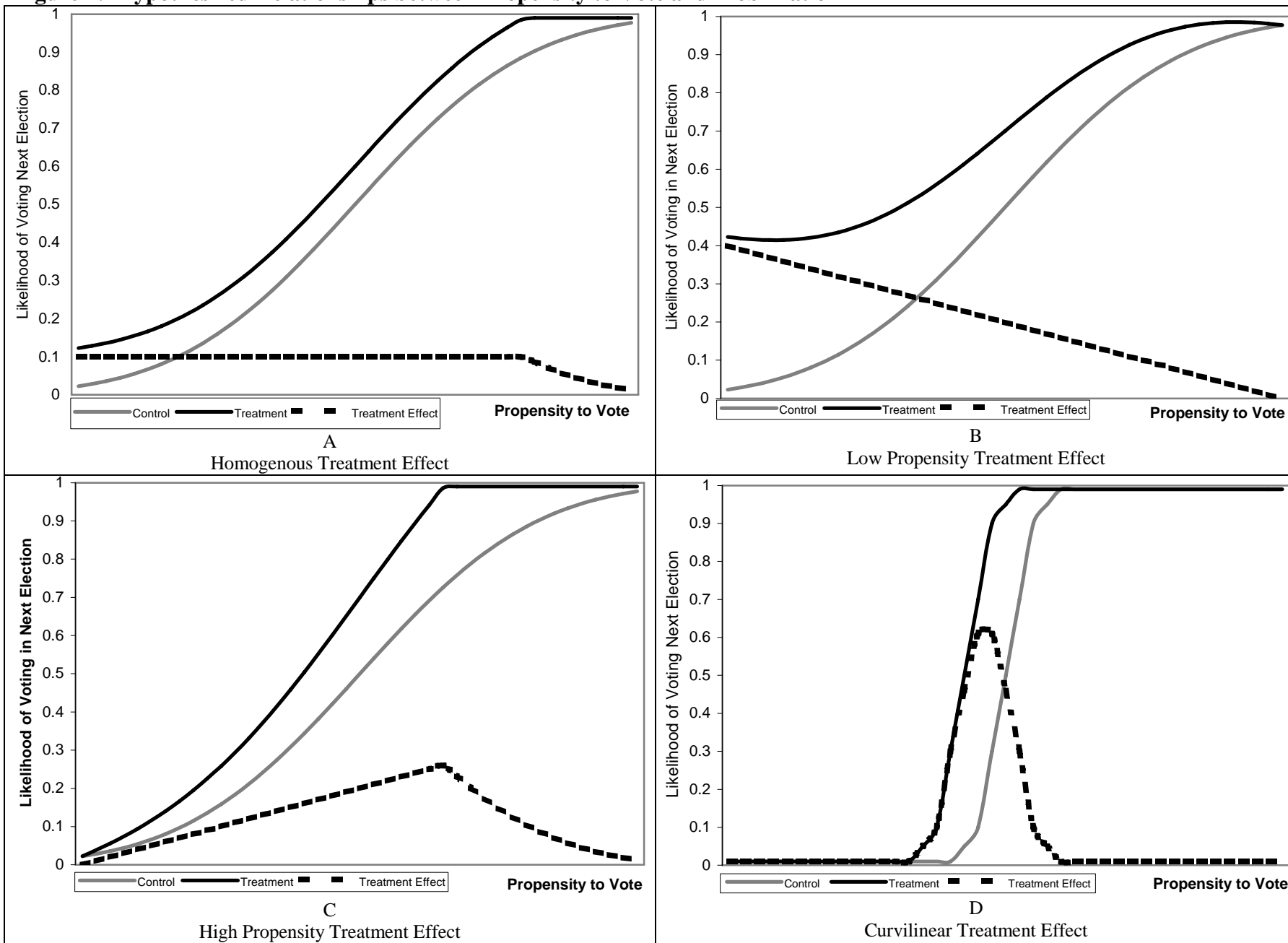
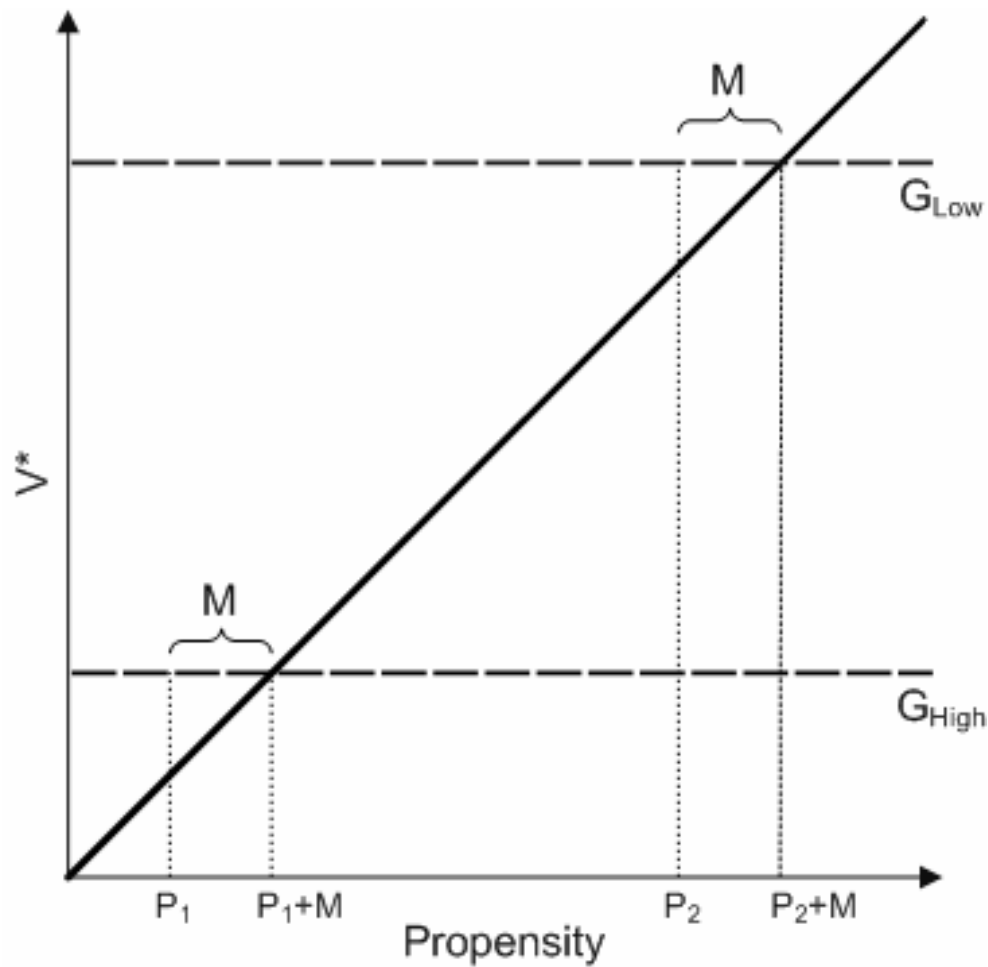


Figure 2: Theoretical Model Predictions



Note: The y-axis, V^* , is the latent likelihood of voting in a particular election. The x-axis, *Propensity*, is a person's 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

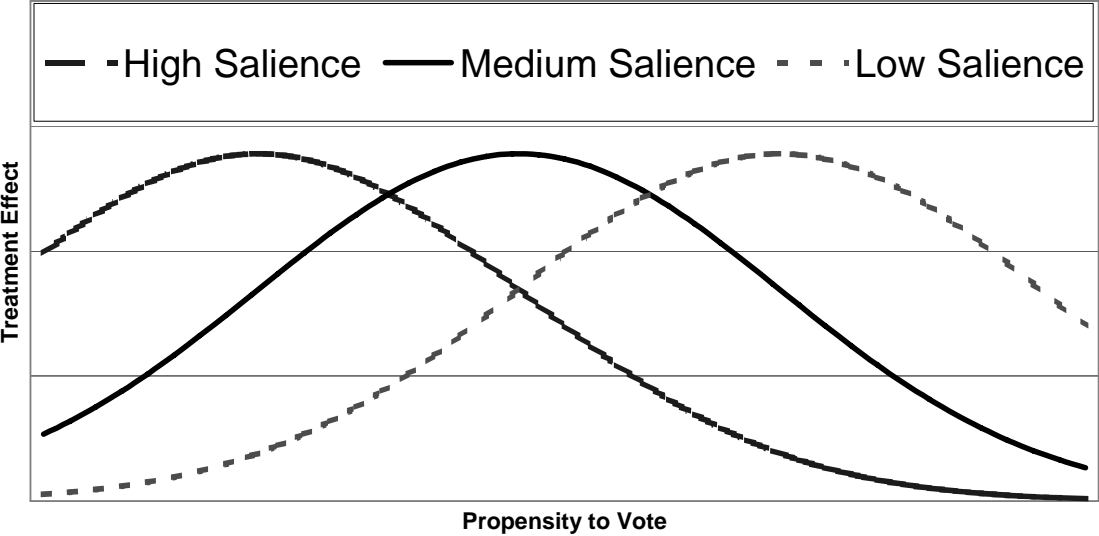
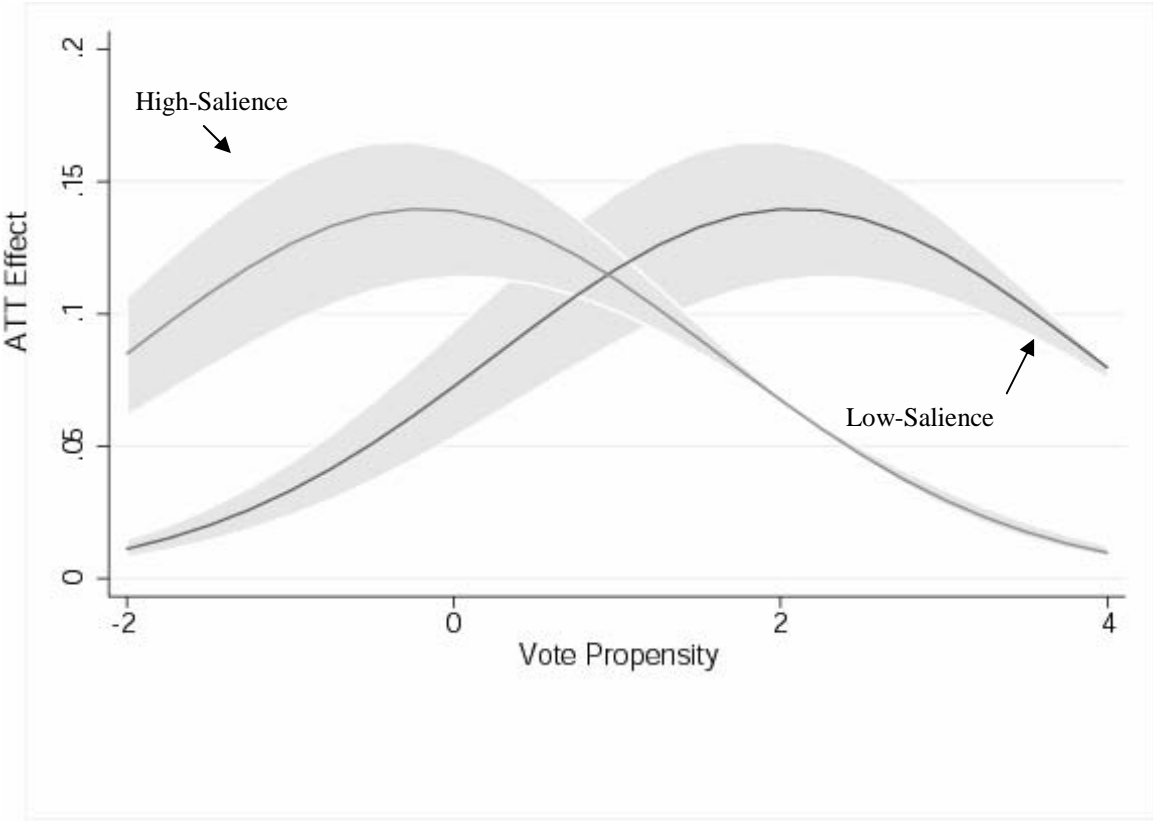


Figure 4: Substantive Effects Generated by Empirical Model



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

Reviewers' Appendix A Descriptions of Experiments

1998 New Haven

The 1998 midterm congressional election in New Haven, CT is the setting of Gerber and Green's (2000) field experiment designed to measure the effectiveness of face-to-face campaign contact (see Table 1, column 1). Every registered voter in the city stood an equal likelihood of being assigned to the face-to-face treatment group.¹⁸ Graduate students were employed to encourage treatment households to vote and successfully contacted 30% of treatment households. As a result of the outreach, the treatment group exhibited turnout 2.7 percentage points higher than the control group.

A subject's propensity to vote was based upon a single prior election, age, and dummies for ward. This constitutes the thinnest vote history on which to base a propensity to vote for all the cities.

2000 Eugene Experiment

During the tightly contested 2000 Presidential race in Oregon, a non-partisan canvassing operation took place in the residential neighborhoods surrounding the University of Oregon in Eugene (see Table 1, column 1). A list of 7,482 registered voters was purchased from a vendor and divided using simple random assignment into a treatment group, receiving knocks on the door in the days immediately prior to Election Day, and a control group, which received no attention from the campaign. The effectiveness of the door-to-door campaign was expected to be somewhat limited since the vote by mail procedure in Oregon allowed individuals to cast ballots two weeks prior to Election Day. In the end, 40 percent of the households in the treatment group were eventually contacted by the campaign and individuals assigned to the treatment group were

¹⁸ Subjects also were assigned to be contacted via phone or mail. We include all subjects in the analysis, but limiting the analysis to only those subjects assigned to the control and face-to-face groups does not meaningfully change the results.

3.4 percentage points more likely to vote than those individuals in the control group. Turnout was uniformly high (69 percent). The propensity score was estimated using vote history, age, party affiliation dummies, year of voter registration, and geography based on zipcode.

2001 Matched Protocol Experiments

The Bridgeport, Columbus, Detroit, Minneapolis, and St. Paul experiments conducted in 2001 used identical protocols (see Green, Gerber, and Nickerson 2003 for more complete descriptions). The Dos Palos protocol differed only in the treatment/control ratio (see Michelson 2003 for a more complete description). Using the official list of registered voters, households were divided into geographically contiguous groupings containing roughly the number of households that could be canvassed by a volunteer in one hour. Within each of these groupings, households were randomly assigned to the treatment (i.e., to be contacted) or control (i.e., no contact) conditions. One individual was randomly selected from the household to be listed on canvassers walk sheets.¹⁹

The actual canvassing of the neighborhoods was nearly identical in execution. Canvassing took place the weekend prior to Election Day and was conducted primarily by young volunteers.²⁰ All of the campaigns were strictly non-partisan; volunteers merely urged individuals to vote in the upcoming election. Canvassers kept track of which households were contacted and closely adhered to the protocol (i.e., only the treatment households were contacted). Thus, the specific volunteers and political context was different for each of the four experiments, but the actual treatment provided was extremely similar.

Bridgeport 2001

¹⁹ This randomly selected person is the only individual from the household used in the analysis.

²⁰ Bridgeport and Dos Palos began canvassing two weeks prior to Election Day.

The Bridgeport experiment was conducted during an uncompetitive school board election where many of the races were uncontested (see Table 1, column 3). The neighborhoods targeted by the campaign were racially mixed and low income (like most of the city). Despite the dearth of homeowners in the sample, the campaign contacted a respectable 28% of subjects in the treatment group. Turnout was extremely low (10%), making the fact that subjects assigned to the treatment group were 3.9 percentage points more likely to vote all the more remarkable.

The propensity score was calculated using vote history from two prior elections, age, year registered, party affiliation, gender, and neighborhood.

Columbus 2001

The Columbus experiment was primarily conducted in the neighborhoods surrounding Ohio State University during a competitive 2001 city council election (see Table 1, column 4). The transient nature of the community caused the contact rate to be lower than would otherwise be expected, 14 percent, however, random assignment assures that the control group contains an equal proportion of persons unable to be contacted. Turnout in the election was low, 9 percent, but those individuals assigned to the treatment group were 1.4 percent points more likely to vote than the control group.

The propensity score was constructed using vote history from three prior elections, party registration, the number of voters in the household, and neighborhood indicators. Registration year and age were not available.

Detroit 2001

The Detroit experiment was conducted during a tightly contested mayoral election in a relatively stable predominantly black neighborhood (see Table 1, column 5). As a result, the contact rate for the canvassing was a respectable 31 percent. Turnout was unusually high for an

off year election (45 percent), and turnout among the treatment group was 2.4 percentage points higher than the control group. The propensity score included three elections, age, registration year, gender, the number of registered voters at the address, and neighborhood.

Minneapolis 2001

The Minneapolis experiment was conducted in two neighborhoods: one populated primarily by University of Minnesota – Twin Cities students, and the other in a low-income minority neighborhood (see Table 1, column 6). The contact rate in both neighborhoods was respectable, but not stellar (19 percent). Most municipal offices were up for grabs, but the headlining race for mayor was not especially close. As a result, overall turnout in the sample was 26 percent. The treatment group was 1.9 percentage points more likely than the control group to vote.

The propensity score was measured using unusually rich Minnesota Voter History file (including 14 elections), age, year of registration, the number of registered voters at the address, and neighborhood.

St. Paul 2001

The experiment in St. Paul was primarily conducted in a stable, working class white neighborhood (but portions of it were also conducted in less affluent minority neighborhoods). The stability of the neighborhood allowed the canvassers to reliably contact individuals and the contact rate was 32 percent (see Table 1, column 7). The mayoral election was decided by a few hundred votes, so 45 percent of the subjects turned out to vote in the election. In the end, the treatment group exhibited a 4.6 percentage point boost in turnout over the control group.

The propensity score was estimated in exactly the same fashion as the Minneapolis 2001 experiment.

Dos Palos 2001

Dos Palos is a small, ethnically mixed community of 4,800 in rural California. Students from California State University – Fresno served as paid canvassers mobilizing voters for a competitive school board election (see Table 1, column 8). Given the high rates of residential stability among registered voters in the area and the diligence of the canvassers, an extremely impressive 72% of subjects assigned to the treatment group were successfully contacted. As a result of the campaign contact, treatment subjects were 3.1 percentage points more likely to vote than control subjects.

The propensity score was calculated using voter history from seven elections, age, year of registration, party affiliation, gender, and geography.

2002 Placebo Controlled Experiments

A highly unusual protocol employing both a placebo and a traditional control group was used during the 2002 Congressional Primaries in Denver and Minneapolis (see Nickerson 2005 for an in depth description of the protocol). Households containing exactly two registered voters were culled from the official voter files. Neighborhoods containing a large number of two voter households were targeted for the experiment. The targeted households were then randomly divided into three groups: treatment, where households received encouragement to vote; placebo, where households received encouragement to recycle; and a control group, where the households received no contact from the campaign whatsoever. Walk lists were drawn up so that volunteers could complete the task in roughly two hours. Canvassers walked from house to house discussing the importance of voting at treatment households, the importance of recycling at placebo households, and not visiting the control households at all. The canvassers flipped between the two scripts with little trouble and carefully recorded who was contacted at the door.

The placebo-controlled design allows for analysis conditional upon answering the door. The rates of turnout for the people who answered the door in the treatment group are compared to the rates of turnout for the people who answered to the door in the placebo group. Since the assignment to treatment and placebo conditions was randomly determined, both sets of people who answered the door should have similar baseline propensities to vote. There should be less noise associated with the estimated treatment effect because the contact rate for the experiment is 100 percent.

The control group can be used to great advantage for calculating propensities to vote in these two experiments. Households were randomly assigned to the control group at the same time the treatment and placebo groups were formed. Therefore, the control group should be perfectly comparable to the other two groups. Since the control group is not involved in calculating the treatment effect, it can be used to calculate the propensity to vote. In this way, we can use the 2002 Congressional Primary to calculate the propensity to vote, thereby obviating the need for the assumption that the propensity to vote in one election correlates highly with the propensity to vote in the next election. In short, the 2002 Denver and Minneapolis experiment constitute the perfect data for the inquiry.

Denver 2002

The overall Denver sample turned out to vote at 39 percent during the 2002 Congressional Primary (see Table 1, column 9). Those reached at the door in the treatment group were 8.6 percentage points more likely to vote than those reached in the placebo group. The Denver experiment is smaller than all the others because the analysis is limited to those who were actually contacted at the door, but the effect size detected is larger because all individuals in the analysis received treatment. The propensity score was calculated using the turnout history

from 7 prior elections, age, gender, registration year, the neighborhood of the respondent, and the 2002 Congressional Primary itself was used as the dependent variable.

Minneapolis 2002

Turnout during the 2002 Congressional Primary in Minneapolis was low (see Table 1, column 10). However, those individuals in the treatment group contacted by the campaign were 10.9 percentage points more likely to vote than the individuals contacted in the placebo group. The propensity score for Minneapolis was calculated using turnout history from 15 prior elections, age, registration year, the neighborhood of the subject, and the 2002 Congressional Primary itself was used as the dependent variable.

2003 Precinct Based Kansas City Experiment

The Kansas City experiment studied a local activist group canvassing on behalf of a public transportation ballot initiative (see Table 1, column 11). The group began canvassing predominately African-American precincts, where it expected to find many supporters of public transportation, four weeks before the election (see Arceneaux 2005 for a more complete description). The campaign canvassed treatment precincts twice and contacted a remarkable 62 percent of the registered voters assigned to treatment precincts. The ballot initiative was not widely viewed as important, so only 31 percent of the subjects turned out to vote. However, the campaign must have passed on some of its enthusiasm for the proposition because the treatment group voted at a rate 4.6 percentage points higher than the control group.

The Kansas City voter file contains an extremely rich voter history, but a paucity of other control variables. As a result the propensity score is based upon 21 elections over a six-year timeframe and dummy variables for precincts. Despite lacking measures of demographic information, it would be surprising if observing the actual voting behavior in 21 previous

elections did not capture much of the information that such covariates impart to the prediction of voting.

Reviewer's Appendix B

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 the two placebo experiments, low contact, and high contact studies. These estimates, reported in Table A1, show that our findings are robust across sample restrictions.

[Table A1 about here]

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.

[Table A2 about here]

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 the 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 A3 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 A3 and Figure A1 about here]

Table A1: Robustness Checks for Probit Model

Variables	Results After Removing:										
	Bridgeport	Columbus	Denver	Dos Palos	Eugene	Kansas City	Minneapolis	Minneapolis Placebo	New Haven	St. Paul	Placebo Experiments
M	0.125 (0.014)	0.127 (0.014)	0.125 (0.014)	0.125 (0.014)	0.129 (0.015)	0.118 (0.016)	0.131 (0.014)	0.125 (0.014)	0.135 (0.016)	0.127 (0.014)	0.123 (0.014)
P	0.565 (0.007)	0.552 (0.007)	0.553 (0.007)	0.561 (0.007)	0.540 (0.007)	0.526 (0.008)	0.539 (0.007)	0.555 (0.007)	0.627 (0.009)	0.542 (0.007)	0.551 (0.007)
G	3.209 (0.049)	3.088 (0.049)	3.243 (0.046)	3.249 (0.047)	3.330 (0.062)	3.326 (0.050)	3.240 (0.047)	3.250 (0.046)	3.372 (0.050)	3.233 (0.046)	3.245 (0.046)
Constant	-1.602 (0.022)	-1.547 (0.023)	-1.620 (0.021)	-1.621 (0.022)	-1.646 (0.025)	-1.671 (0.024)	-1.621 (0.022)	-1.622 (0.021)	-1.666 (0.022)	-1.618 (0.021)	-1.620 (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

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

Table A2: Testing for the Parabolic-Shape of the Relationship Between the Treatment Effect and Voting Propensity with a Linear Probability Model

	Full Sample	Results After Excluding:								
		Bridgeport	Columbus	Denver	Detroit	Dos Palos	Eugene	Kansas City	Minneapolis	Minneapolis Placebo
M	0.112 (0.014)	0.113 (0.015)	0.105 (0.014)	0.108 (0.014)	0.115 (0.015)	0.119 (0.016)	0.117 (0.015)	0.130 (0.022)	0.114 (0.014)	0.110 (0.014)
P	0.192 (0.004)	0.197 (0.004)	0.202 (0.004)	0.192 (0.004)	0.188 (0.005)	0.198 (0.004)	0.189 (0.004)	0.185 (0.005)	0.189 (0.004)	0.193 (0.004)
P ²	-0.019 (0.003)	-0.019 (0.003)	-0.027 (0.003)	-0.019 (0.003)	-0.011 (0.003)	-0.021 (0.003)	-0.019 (0.003)	-0.027 (0.004)	-0.023 (0.003)	-0.020 (0.003)
M*P	0.028 (0.015)	0.029 (0.015)	0.028 (0.015)	0.028 (0.015)	0.036 (0.016)	0.028 (0.017)	0.040 (0.016)	0.018 (0.018)	0.026 (0.015)	0.027 (0.015)
M*P ²	-0.015 (0.010)	-0.017 (0.010)	-0.013 (0.010)	-0.014 (0.010)	-0.018 (0.010)	-0.017 (0.011)	-0.015 (0.010)	-0.014 (0.015)	-0.016 (0.010)	-0.014 (0.010)
Constant	0.390 (0.004)	0.402 (0.005)	0.418 (0.005)	0.391 (0.004)	0.373 (0.005)	0.400 (0.005)	0.348 (0.005)	0.421 (0.005)	0.403 (0.005)	0.394 (0.004)
Observations	41,863	40,059	39,385	41,301	36,909	39,888	36,806	32,151	39,037	41,469
F	1,417.512	1,419.966	1,388.028	1,373.120	1,312.326	1,389.086	1,282.738	929.328	1,234.924	1,401.331
Adjusted R ²	0.150	0.150	0.150	0.140	0.150	0.150	0.150	0.130	0.140	0.150

Table A2 Continued

	Results After Excluding:		
	New Haven	St Paul	Studies with Limited Vote History
M	0.102 (0.015)	0.111 (0.015)	0.103 (0.015)
P	0.197 (0.005)	0.186 (0.004)	0.203 (0.005)
P ²	-0.007 (0.003)	-0.018 (0.003)	-0.006 (0.003)
M*P	0.019 (0.015)	0.031 (0.015)	0.019 (0.015)
M*P ²	-0.013 (0.010)	-0.015 (0.010)	-0.015 (0.010)
Constant	0.347 (0.005)	0.389 (0.005)	0.361 (0.005)
Observations	31,963	39,662	30,159
F	1,271.866	1,258.567	1,278.035
Adjusted R ²	0.170	0.140	0.180

Table A3: Including *G* in the Linear Probability Model

	Results After Excluding:									
	Full	Bridgeport	Columbus	Denver	Detroit	Dos Palos	Eugene	Kansas City	Minneapolis	Minneapolis Placebo
M	0.079 (0.038)	0.068 (0.039)	0.077 (0.038)	0.077 (0.038)	0.082 (0.037)	0.102 (0.043)	0.043 (0.053)	0.104 (0.052)	0.082 (0.038)	0.066 (0.037)
P	0.130 (0.010)	0.150 (0.011)	0.167 (0.012)	0.130 (0.010)	0.129 (0.010)	0.142 (0.011)	0.117 (0.011)	0.100 (0.011)	0.117 (0.010)	0.133 (0.009)
P ²	0.050 (0.007)	0.060 (0.007)	0.033 (0.009)	0.050 (0.007)	0.047 (0.007)	0.047 (0.007)	0.078 (0.008)	0.046 (0.007)	0.047 (0.007)	0.048 (0.006)
G	1.210 (0.025)	1.240 (0.028)	1.182 (0.031)	1.212 (0.025)	1.179 (0.025)	1.203 (0.026)	1.302 (0.034)	1.217 (0.027)	1.207 (0.026)	1.200 (0.024)
M*P	0.087 (0.036)	0.087 (0.038)	0.089 (0.034)	0.086 (0.036)	0.090 (0.035)	0.104 (0.046)	0.056 (0.052)	0.086 (0.044)	0.082 (0.036)	0.081 (0.036)
M*P ²	-0.0003 (0.026)	0.003 (0.027)	0.005 (0.027)	0.000 (0.026)	-0.003 (0.025)	-0.015 (0.031)	0.003 (0.035)	0.008 (0.035)	-0.002 (0.026)	0.004 (0.025)
M*G	0.088 (0.096)	0.115 (0.100)	0.095 (0.098)	0.084 (0.096)	0.092 (0.096)	0.045 (0.104)	0.204 (0.160)	0.046 (0.111)	0.082 (0.097)	0.115 (0.096)
P*G	0.165 (0.022)	0.125 (0.024)	0.090 (0.027)	0.165 (0.022)	0.160 (0.022)	0.144 (0.023)	0.216 (0.030)	0.212 (0.023)	0.188 (0.022)	0.159 (0.021)
P ² *G	-0.207 (0.017)	-0.234 (0.019)	-0.171 (0.022)	-0.208 (0.017)	-0.178 (0.018)	-0.203 (0.018)	-0.314 (0.022)	-0.207 (0.018)	-0.202 (0.018)	-0.204 (0.017)
M*P*G	-0.152 (0.086)	-0.151 (0.090)	-0.154 (0.085)	-0.151 (0.086)	-0.140 (0.086)	-0.181 (0.102)	-0.050 (0.154)	-0.148 (0.095)	-0.142 (0.087)	-0.140 (0.086)
M*P ² *G	-0.046 (0.068)	-0.055 (0.071)	-0.058 (0.071)	-0.043 (0.068)	-0.049 (0.068)	-0.018 (0.076)	-0.052 (0.108)	-0.056 (0.077)	-0.041 (0.069)	-0.054 (0.068)
Constant	-0.056 (0.010)	-0.068 (0.011)	-0.043 (0.013)	-0.055 (0.010)	-0.053 (0.010)	-0.052 (0.011)	-0.080 (0.012)	-0.057 (0.012)	-0.055 (0.011)	-0.051 (0.010)
Observations	41,863	40,059	39,385	41,301	36,909	39,888	36,806	32,151	39,037	41,469
F	1,350.546	1,268.344	1,159.947	1,328.439	1,311.072	1,301.370	999.102	1,054.524	1,215.246	1,329.523
Adjusted R2	0.260	0.260	0.250	0.260	0.280	0.270	0.230	0.270	0.260	0.260

Table A3 Continued

	Results After Excluding:		
	New Haven	St Paul	Studies with Limited Vote History
M	0.084 (0.037)	0.078 (0.038)	0.073 (0.039)
P	0.129 (0.009)	0.122 (0.010)	0.149 (0.010)
P ²	0.035 (0.006)	0.052 (0.007)	0.045 (0.007)
G	1.145 (0.027)	1.215 (0.025)	1.176 (0.029)
M*P	0.090 (0.034)	0.090 (0.036)	0.090 (0.036)
M*P ²	-0.001 (0.025)	-0.0004 (0.026)	0.002 (0.026)
M*G	0.051 (0.098)	0.088 (0.096)	0.080 (0.102)
P*G	0.191 (0.022)	0.170 (0.022)	0.152 (0.024)
P ² *G	-0.136 (0.019)	-0.212 (0.017)	-0.164 (0.020)
M*P*G	-0.192 (0.085)	-0.152 (0.086)	-0.192 (0.089)
M*P ² *G	-0.041 (0.070)	-0.046 (0.068)	-0.052 (0.072)
Constant	-0.042 (0.010)	-0.059 (0.010)	-0.053 (0.011)
Observations	31,963	39,662	30,159
F	1,257.673	1,272.647	1,181.446
Adjusted R ²	0.300	0.260	0.300

Figure A1: Linear Probability Model Substantive Effects

