

The Journal of Politics

Increasing Inequality: The Effect of GOTV Mobilization on the Composition of the Electorate --Manuscript Draft--

Manuscript Number:	JOP-D-13-00071R2
Full Title:	Increasing Inequality: The Effect of GOTV Mobilization on the Composition of the Electorate
Article Type:	Article
Section/Category:	American Politics - Behavior
Corresponding Author:	Anthony Fowler, Ph.D. University of Chicago Chicago, IL UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Chicago
Corresponding Author's Secondary Institution:	
First Author:	Ryan D Enos, Ph.D.
First Author Secondary Information:	
Order of Authors:	Ryan D Enos, Ph.D. Anthony Fowler, Ph.D. Lynn Vavreck, Ph.D.
Order of Authors Secondary Information:	
Abstract:	Numerous get-out-the-vote (GOTV) interventions are successful in raising voter turnout. However, these increases may not be evenly distributed across the electorate and could potentially increase the differences between voters and non-voters. By analyzing individual level-data, we reassess previous GOTV experiments to determine which interventions mobilize under-represented citizens versus those who regularly turn out. We develop a generalized and exportable test which indicates whether a particular intervention reduces or exacerbates disparities in political participation and apply it to 24 previous experimental interventions. On average, current mobilization strategies significantly widen disparities in participation by mobilizing high-propensity individuals more than the under-represented, low-propensity citizens. The results hold troubling implications for the study and improvement of political inequality, but the methodological procedures laid out in this study may assist the development and testing of future strategies which reverse this pattern.

Increasing Inequality:
The Effect of GOTV Mobilization on the Composition of the Electorate

Ryan D. Enos
Department of Government
Harvard University
1737 Cambridge Street, K406
Cambridge, MA 02138
renos@gov.harvard.edu

Anthony Fowler
Harris School of Public Policy Studies
University of Chicago
1155 East 60th Street, 165
Chicago, IL 60637
anthony.fowler@uchicago.edu

Lynn Vavreck
Departments of Political Science
University of California, Los Angeles
4289 Bunche Hall
Los Angeles, CA 90095
lvavreck@ucla.edu

Footnote: Author order is alphabetical. An online appendix for this article is available at [www.cambridge.org/cjo/\[TBD\]](http://www.cambridge.org/cjo/[TBD]) containing more details on the data and analyses and additional results. Data and supporting materials necessary to reproduce the numerical results and figures in the article are currently available at www.dropbox.com/sh/b6257dcqkr27ald/ZuQfx-KUzc.

Abstract

Numerous get-out-the-vote (GOTV) interventions are successful in raising voter turnout. However, these increases may not be evenly distributed across the electorate and could potentially increase the differences between voters and non-voters. By analyzing individual level-data, we reassess previous GOTV experiments to determine which interventions mobilize under-represented citizens versus those who regularly turn out. We develop a generalized and exportable test which indicates whether a particular intervention reduces or exacerbates disparities in political participation and apply it to 24 previous experimental interventions. On average, current mobilization strategies significantly widen disparities in participation by mobilizing high-propensity individuals more than the under-represented, low-propensity citizens. The results hold troubling implications for the study and improvement of political inequality, but the methodological procedures laid out in this study may assist the development and testing of future strategies which reverse this pattern.

Keywords: turnout; participation; inequality; experiments; get-out-the-vote

Scholars are increasingly interested in inequalities in political participation and their consequences for political outcomes (APSA Task Force 2004; Bartels 2008, 2009; Dahl 2006; Gilens 2005, 2012; Schlozman, Verba, and Brady 2012). At the same time, political scientists increasingly use large-scale field experiments to test methods for increasing political participation. The findings of these experiments have been adopted by political campaigns and are an increasingly important feature of electoral politics (Gerber et al. 2011; Green and Gerber 2008; Issenberg 2010, 2012). Often (but not always), a goal of these get-out-the-vote (GOTV) strategies is to reduce the participation gap – the extent to which the electorate differs from the voting-eligible population. While many experiments successfully increase average levels of voter participation, they may not affect all citizens equally. Instead, an experiment may affect the types of citizens already represented in the political process more than underrepresented citizens. In this article, we explicitly test whether GOTV treatments tend to reduce or exacerbate the gap in political participation. We find that GOTV interventions, on average, tend to magnify the participation gap. This finding is widely important for the study and equalization of political representation, political campaigns, and voting behavior.

In a typical GOTV experiment, a researcher will randomly assign individuals, households, or geographic regions to receive a particular treatment. Then, the researcher will typically collect voting data from public records and estimate the average effect of the treatment on voter turnout by calculating a simple difference-in-means. To our knowledge, there are four primary motivations for conducting GOTV experiments. First, scholars and civic groups may want to increase political participation as an end in and of itself. Second, field experiments can improve the efficiency of political campaigns. Third, these experiments might inform the study of voting behavior and human behavior in general. Fourth, the participation gap may hold critical political and policy outcomes, and GOTV experiments can identify methods for reducing this gap. For all but the first motivation, the traditional GOTV analysis of calculating average treatment effects is usually insufficient. In addition to knowing *how many* voters were mobilized by a particular treatment, we would also like to know *which types* of

voters were mobilized. We offer a generalized method for assessing one particular type of heterogeneity in experimental treatment effects which can lend insights for all three of the preceding motivations, although we focus mainly on the last.

Because voters are systematically unrepresentative of the rest of the population, some individuals and groups may be underrepresented relative to others (Lijphart 1997; Verba, Scholzman, and Brady 1995). By studying the determinants of voter turnout, scholars may be able to propose policies to reduce differences in political participation. We develop a single statistical procedure that can be applied to any GOTV experiment to explicitly test whether the intervention reduced or exacerbated disparities in voter turnout. The test can be easily implemented for any previous experiment, often without the collection of any additional data. We apply this method to 24 experimental interventions from 11 published papers to assess the overall implications of the get-out-the-vote paradigm for equality and representation. Interpreting the point estimates directly, 16 of the interventions are found to have exacerbated the participation gap while only 8 are found to have reduced it. Moreover, 8 of the exacerbating effects are statistically significant while only 2 of the reducing effects cross this threshold. A pooled analysis of all interventions reveals a large and statistically significant exacerbating effect of GOTV interventions, on average, on disparities in political participation.

By analyzing many experiments at once, we build upon the work of Arceneaux and Nickerson (2009) who re-analyze 11 previous door-to-door canvassing experiments to determine which subset of citizens should be targeted to achieve the most efficient allocation of campaign resources. They argue that campaigns should target higher-propensity voters in low-salience elections and lower-propensity voters in high-salience elections. In this article, we expand the number of interventions to 24, examine many different types of mobilization strategies, and move beyond the focus of campaigns and electioneering to assess the effects of voter mobilization on inequality and by extension political representation. Some of our subsequent analyses are consistent with the results of Arceneaux and Nickerson. For example, the ease with which low-propensity individuals can be mobilized increases

with the salience of the election (shown in the Online Appendix). However, in the typical election, high-propensity citizens are much easier to mobilize and this remains true even in many high-salience elections. As a result, modern mobilization strategies, which are being actively used by interest groups and campaigns, can dramatically and systematically change the composition of the electorate.[1]

Get-Out-the-Vote and the Political Consequences of the Participation Gap

“The existence of political equality is the fundamental premise of democracy” (Dahl 2006). Dahl’s assertion presents a difficult challenge for democratic societies because political equality is undermined by disparities in voter participation. As Verba, Schlozman, and Brady (1995) put it: “since democracy implies not only government responsiveness to citizen interests but also equal consideration of the interests of each citizen, democratic participation must also be equal....No democratic nation – certainly not the United States – lives up to the ideal of participatory equality.” As Citrin, Schickler, and Sides (2003) explain, “there are meaningful differences between voters and non-voters,” and “tiny changes in the distribution of the popular vote (and thus small differences in turnout) can make an enormous difference to the nation’s politics” (p. 88). Scholars have shown that increased turnout is associated with more equitable municipal spending (Hajnal and Trounstein 2005), altered election results (Hansford and Gomez 2010), less interest group capture (Anzia 2011), increased working class representation (Fowler 2013), and general legislative responsiveness (Griffin and Newman 2005).

Recognizing the importance of participation to a healthy democracy and electoral outcomes, political scientists have long studied the determinants of political participation (e.g. Putnam 2000; Wolfinger and Rosenstone 1980; Verba, Schlozman, and Brady 1995). The introduction of field experiments gives political scientists another tool for understanding how campaigns and policy-makers can decrease inequalities in turnout, and scholars note the potential for field experiments to increase participation among underrepresented groups. For example, Avery (1989) cites door-to-door canvassing as a method to increase participation among regular non-voters; Michelson, Garcia Bedolla,

and Green (2008) point to a “multi-year effort to increase voting rates among infrequent voters” that relied entirely on GOTV experiments; and Green and Michelson (2009) argue that GOTV campaigns can alter the “age, socioeconomic, and racial/ethnic disparities between voters and nonvoters” (see also Garcia Bedolla and Michelson 2012).

What has not been acknowledged by practitioners of voter mobilization is that GOTV methods may actually exacerbate the differences between voters and the voting-eligible population. Furthermore, no previous study has explicitly assessed the effects of voter mobilization on the participation gap and representational inequality. Given the prevalence of GOTV field experiments and the fact that political campaigns are now adopting the methods that these experiments have shown to be effective (Gerber et al. 2011; Green and Gerber 2008; Issenberg 2010, 2012), an assessment is warranted.

How Can GOTV Treatments Exacerbate the Participation Gap?

For both theoretical and empirical reasons, our claim that get-out-the-vote treatments may exacerbate the participation gap may be counterintuitive. First, many scholars assume that increased turnout is unambiguously good for democracy and equality (Key 1949; Lijphart 1997; Schattschneider 1960). For example, Schattschneider writes “Unquestionably, the addition of forty million voters (or any major fraction of them) would make a tremendous difference [in the quality of representation]” (1960, p. 101). Schattschneider’s observation is obviously true in the extreme because with universal turnout there would be no difference between the population of voters and the voting-eligible population. However, when not everyone votes regularly, especially in systems with widespread non-participation like the United States, a marginal increase in voter turnout may actually exacerbate differences if the increase is concentrated among high-propensity voters, the types of citizens who are voting at high rates anyway.

To illustrate this possibility, Figure 1 shows three hypothetical experimental treatments. For all three treatments and a control group, the lines plot voter turnout across varying underlying propensities to vote. The underlying propensity to vote can be thought of as the probability that the voter would have voted absent a treatment. All three treatments have the same average effect. For an individual with an average propensity, the treatments all increase the probability of voting by 10 percentage points. Despite this large, positive average effect, the treatments have starkly different implications for equality of participation. One treatment effect (blue) is concentrated among low-propensity voters and therefore reduces the participation gap. Another treatment (red) is concentrated among high-propensity voters, thereby exacerbating differences. The other treatment effect (green) is homogeneous and is therefore neutral in regards to these disparities.

[Figure 1]

Our findings may also be counterintuitive because at high-propensity levels we expect to see a ceiling effect. If an individual is sure to vote in the absence of a treatment, then no treatment can exhibit a positive effect. For this reason, we might expect most treatments to naturally favor equality. After all, low-propensity voters have more room to increase their turnout probabilities because they are voting at lower rates. These factors make our subsequent results all the more surprising. Despite the possibility of a ceiling effect and despite the large numbers of low-propensity citizens that can be mobilized, we still find that most GOTV interventions *exacerbate* the already stark disparities in voter turnout.

Previous Studies of “Who is Mobilized?”

The varying treatment effects across different underlying propensities to vote that are illustrated in Figure 1 can be thought of as interactions between the propensity to vote and the treatment. Some previous GOTV studies have examined interactive effects in attempts to test mechanisms or to look for stronger or weaker effects across subgroups. For example, Nickerson and Rogers (2010) show that

discussing a “voting plan” tends to mobilize individuals who live alone but has little effect for others; Alvarez, Hopkins, and Sinclair (2010) find that partisan campaign contacts are most effective in mobilizing new registrants; Gerber and Green (2000b) find that non-partisan leaflets are most effective among non-partisans who have recently voted; and Green and Gerber (2008) find that canvassing typically boosts turnout among those who voted in the previous election.

Despite these examples, this type of analysis is rare. Experimental researchers do not typically explore all possible interactive effects of their treatment – and rightfully so, because atheoretical testing of multiple interactive effects will generate false positive findings (Gabler et al. 2009; Pocock et al. 2002). While researchers could reduce this problem by testing for interactions with split samples (Green and Kern 2012), the atheoretical testing of many hypotheses is not recommended. This methodological issue poses a problem for those interested in the participation gap and the related effects on representation and policy outcomes if this interest is not factored into the design phase of an experiment. As previously discussed, the typical GOTV study simply tells us how many people were mobilized by a particular treatment but not which types of individuals were mobilized. To learn who is mobilized, an interactive test is called for, but this increases the danger of producing false positives.

An Empirical Test for the Effect of a Treatment on the Participation Gap

We develop a single test which explicitly assesses whether an experimental treatment exacerbates or reduces disparities in participation. We propose a procedure which reduces an individual’s pre-treatment characteristics onto a single dimension, her propensity to vote. Having estimated this propensity, we test whether the experimental treatment effect varies across this single variable.

Many demographic factors predict an individual’s propensity to vote. In our analysis we reduce all of the factors onto one scale - a propensity score - that indicates the *a priori* probability that a person with those characteristics will vote in the absence of any GOTV intervention. Of course, not all

relevant demographic variables are available through public records used in GOTV experiments. Voter files typically indicate an individual's age, gender, geographic location, and previous turnout history. Variables like race and party registration are only available in certain states, and personal information such as income, education, or church attendance is never available. Even the limited variables that are available serve as a proxy for the extent to which a particular individual may be represented in the political process. For the purposes of this article, we do not care *why* these demographic variables predict voter turnout. The fact that any variables can systematically predict the propensity to vote suggests that there are meaningful disparities in political participation, and our test assesses whether a particular treatment reduces or exacerbates those disparities. Later in the article, we employ survey data to show that this propensity variable is strongly correlated with demographic factors that portend specific policy positions such as income, education, church attendance, and marital status as well as political attitudes on taxes, minimum wage, federal spending, affirmative action, and other important issues.

If a specific experimental treatment mobilizes low-propensity citizens more than high-propensity citizens, then we will say that the treatment has reduced the participation gap. In other words, the demographic gap between voters and the greater population has been reduced. However, if a treatment mobilizes high-propensity citizens more than low-propensity citizens, we will say that the intervention has exacerbated the participation gap. In this scenario, the voting population has become even more unrepresentative of the general population.

Our estimation procedure involves three specific steps. First, we estimate a propensity score for every individual in the sample by regressing voter turnout on every available demographic variable for each individual in the control group.[2] The specification for estimation should be as flexible as possible given the amount of available data. We restrict this part of the analysis to subjects assigned to the control group because we want to estimate the propensity of each individual in the absence of any treatment. Because individuals have been randomly assigned, we know that the propensity scores of

the control group are representative of the treatment group as well. Then, for every individual in the sample, we calculate their predicted probability of voting. For each individual, this score indicates their *a priori* probability of voting in the absence of a treatment. This score represents our single propensity variable that we employ to assess the effect of each treatment on differences in participation. For the second step, we rescale the propensity variable such that the mean equals zero and the standard deviation equals one. This is done by subtracting the mean from each individual score and then dividing by the standard deviation. This step allows us to reasonably compare treatments across different types of elections and populations, and also improves the interpretation of our subsequent estimates. Lastly, we estimate the following interactive model by OLS to test whether the treatment effect increases or decreases as the propensity score increases:

$$(1) \quad \text{Turnout}_i = \alpha + \beta * \text{Treatment}_i + \gamma * \text{Propensity}_i + \delta * \text{Treatment}_i * \text{Propensity}_i + \varepsilon.$$

The two coefficients of interest are β and δ . The *treatment* coefficient, β , represents the treatment effect for an individual with an average propensity score (this is not necessarily the same as the average treatment effect). The interactive or multiplicative coefficient, δ , represents the extent to which the treatment effect varies with the propensity score. For example, a δ of .01 indicates that the treatment effect increases by 1 percentage point, on average, for every standard deviation increase in propensity. If δ is greater than zero, then the treatment has exacerbated the participation gap, while a δ less than zero indicates that the treatment has reduced the gap. If the interactive effect is non-linear, then we should not interpret the model literally by imputing the treatment effect at various levels of propensity. Nonetheless, even relaxing any assumption of linearity, the sign of δ still indicates whether the treatment is on average greater or smaller for high-propensity citizens.

Two tricky methodological issues could theoretically pose a problem for our analysis but do not. First, the propensity variable is estimated from one regression and incorporated into a second regression. As a result, the second regression may yield incorrect standard errors by failing to

incorporate the uncertainty from the first regression. We can assess this additional uncertainty with a non-parametric bootstrap, and in each case in our data the bootstrapped standard errors are virtually identical to those where we ignore the initial uncertainty of the propensity variable. Second, we might worry that our initial regression “over-fits” the data by fitting random noise rather than an underlying relationship. If this were the case, our subsequent estimates could be biased. Specifically, if we over-fit the data in the first regression, that would lead to a downward bias in the interaction term in the second regression because our propensity variable would not predict turnout out-of-sample as well as it does within-sample. We can address this issue by randomly partitioning the control group into two groups, estimating the propensity variable with one group, and then running the second regression using only the second group. This procedure also produces results nearly identical to the original setup.[3]

Having estimated this parametric model, we also explore the non-linearity of the interactive effect through kernel regressions. By plotting the probability of turnout as a non-parametric function of the propensity score for both the treated and control observations, we obtain a more detailed picture of how the treatment effect varies across different propensity scores.

This analysis provides a single, theoretically-based method for evaluating the effect of a treatment on participation differences. In the following sections, we apply this procedure to 24 different GOTV interventions and then pool all of the data to assess the overall effect of these programs on the participation gap.

A Detailed Example: The “Neighbors” Treatment

To clarify our procedure, we will discuss one experiment; Gerber, Green, and Larimer’s (2008) Neighbors treatment; in greater detail. Following this discussion, we will provide a broader analysis of the others. In this experiment, the authors randomly assigned registered voters to receive postcards in the run up to a low-salience 2006 primary election in Michigan. The sample included registered voters in Michigan who voted in the 2004 general election. Subjects in the control condition received no

contact at all, while subjects in the treatment condition received a memorable postcard, indicating the previous turnout behavior of individuals in their neighborhood. All at once, this powerful treatment informed citizens about the upcoming election, indicated that a researcher was watching them, and threatened to share their turnout behavior with others in the neighborhood.

The Neighbors treatment had a massive average treatment effect, raising turnout by 8 percentage points. However, despite a strong average effect, the treatment may not be equally effective for all types of citizens. For the purposes of our study, we want to know whether this treatment tends to reduce or exacerbate the participation gap by mobilizing low or high-propensity citizens.

To estimate our propensity variable for each subject in the experiment, we utilize all the pre-treatment variables available in the data provided by the authors. We know whether each subject voted in the primary and general elections of 2000, 2002, and 2004; the gender of each subject; their age, and their household size (calculated as the number of voters registered at the same address). Restricting our analysis to the control group, we divided subjects' ages into 13 categories and then regressed the dependent variable (turnout in a 2006 primary) on turnout in each of the 6 previous elections, gender, age category fixed effects, and household size fixed effects. We could have chosen a more flexible model with interactive effects, but we must be careful to avoid over-fitting. Having estimated the model by OLS, we predict for each individual in both the control and treatment groups their *a priori* probability of voting. These predicted probabilities have a mean of .30 and standard deviation of .13, which we then rescale so that the mean is 0 and the standard deviation is 1. The r-squared of .21 indicates that our model explains a significant proportion but not all of the variation in turnout.[4]

Having estimated the propensity variable, we test whether the treatment effect varies across different propensities. As described earlier, we regress turnout on the treatment, propensity, and the interaction of the two. The coefficient on the treatment variable is .080 with a standard error of .003 (clustered by household), indicating that the treatment effect was 8 percentage points for the hypothetical individual with an average propensity score.[5] This is remarkably close to the average

treatment effect reported by Gerber, Green, and Larimer, but this need not be the case. The coefficient on the interaction term is of greatest interest because it indicates how the treatment effect changes as propensity changes. Here, we estimate a coefficient of .016 with a standard error of .003 ($p < .01$), indicating that on average the treatment effect increases by 1.6 percentage points for every standard deviation increase in propensity. According to our statistical test, the “neighbors” treatment exacerbated the participation gap despite significantly raising the average level of turnout.

One way to assess the substantive size of this interactive effect is to compare the treatment effects for two hypothetical groups whose propensity scores are respectively two standard deviations above and below the mean. According to our model, the conditional average treatment effect of the “neighbors” intervention was 11.2 percentage points ($.080 + 2*.016 = .112$) for the high-propensity group, while the effect was only 4.8 percentage points for the low-propensity group. As such, the *neighbors* treatment effect is more than 2 times greater for the highest propensity individuals in the sample compared to the lowest propensity individuals. This difference is substantively large and statistically significant, so we conclude that this experimental intervention significantly exacerbates the participation gap.

Having conducted our formal statistical test, we take a closer look at this interactive effect with a graphical, nonparametric approach. In Figure 2, we employ kernel regressions to plot the probability of turnout across individuals’ *a priori* probability of voting. We plot these relationships separately for both the control condition and the treatment condition. We display these curves in the top panel of Figure 2. For low-propensity voters, there is little difference between the curves, indicating a small conditional average treatment effect. However, as propensities increase, the gap between the groups increases as well, indicating that the treatment effect is becoming stronger for high-propensity citizens. In the bottom panel of Figure 2, we plot the difference between these two kernel regressions, showing how the treatment effect varies across different propensities. In this panel, the higher the curve, the greater the treatment effect, so a positively sloping curve indicates that the treatment effect increases as

propensity increases. The curve in the bottom panel turns downward at the highest propensities, demonstrating the expected ceiling effect. However, prior to this downward turn, the curve is positive and increasing, indicating that the treatment has an increasingly strong effect as the underlying propensity to vote increases – thereby exacerbating existing differences in voter turnout.

[Figure 2]

Later, we apply these same procedures to 24 different experimental treatments. This allows us to statistically assess the effects of GOTV treatments on the participation gap and also to visualize the way in which these effects vary across different experiments. The Neighbors treatment is not an anomaly: numerous experimental interventions exacerbate the participation gap while few interventions significantly reduce it. First however, we establish that the propensity score measure we have described is associated with politically meaningful variation in the electorate.

Does the propensity score capture meaningful differences between voters?

We may not care whether GOTV interventions mobilize high or low-propensity citizens if our propensity variable does not capture meaningful characteristics of the electorate that are associated with policy preferences. Political scientists have extensively studied the correlates of turnout (e.g. Putnam 2000; Wolfinger and Rosenstone 1980; Verba, Schlozman, and Brady 1995), and this literature suggests that high-propensity citizens will be systematically different from low-propensity citizens across a number of politically-relevant variables. Even though we only have data on vote history and a few demographics, we argue that our propensity variable is a proxy for many characteristics that we care about such as socio-economic status and issue positions.

We test the political relevance of our propensity variable using survey data. We employ data from the 2008 versions of the Cooperative Congressional Election Study [6] (CCES) and the Cooperative Campaign Analysis Project [7] (CCAP). For each survey respondent, we generate our propensity variable, using turnout in the 2008 general election as the dependent variable and only using

information that would be available in public records as the independent variables: age, race, gender, household size, state, party registration, and vote history in previous elections. To keep in line with typical GOTV samples, we only include registered voters in the samples. Also, because survey respondents were matched to statewide voter files, we can use validated turnout data instead of reported behavior in the analysis.

Having generated our propensity variable using only those variables available in public records, we test whether this variable captures other meaningful features of the citizenry. Table 1 reports the results of 33 separate regressions. In each case, we regress a demographic characteristic or political attitude on the propensity variable. The coefficient indicates the extent to which the characteristic changes, on average, for every standard deviation increase in propensity. With the exceptions of family income and party identification, all dependent variables are coded as dummies, so the coefficients can be interpreted as changes in probability. For example, a single standard deviation increase in propensity corresponds with a \$6,000 increase in family income, a 6 percentage point increase in the probability of a college degree, and a 3 percentage point increase in the probability of approving of George W. Bush. To summarize, high-propensity citizens, as identified by our method, are wealthier, more educated, more likely to attend church, more likely to be employed, more likely to approve of Bush, more conservative, and more Republican. They are more supportive of abortion rights and less supportive of withdrawing troops from Iraq, domestic spending, affirmative action, minimum wage, gay marriage, federal housing assistance, and taxes on wealthy families.

[Table 1]

Even though our propensity variable relies on a small number of sparse measures that are readily available from public records, it corresponds strongly with numerous demographic characteristics and issue positions which are highly relevant in American politics. As a result, if GOTV interventions tend to mobilize high-propensity citizens over those with low propensities, they will make the electorate wealthier, more educated, more religious, and more conservative on a number of

important issues. However, high-propensity citizens are not more conservative on all issues. In fact, they are more liberal on abortion and they are no different on healthcare and immigration. The abortion result is consistent with previous findings that high socioeconomic-status citizens are more supportive of abortion rights (Bartels 2008). The propensity variable is also highly correlated with intensity of preferences. For example, high-propensity citizens are much more likely to have an extreme ideology or a strong party identification. As a result, GOTV interventions which mobilize high propensity citizens are likely to increase the polarization of the electorate in addition to changing its demographic composition. This analysis demonstrates that our propensity variable captures meaningful political differences. Having demonstrated the wealth of information captured by our propensity variable, we apply our test to experimental data.

Applying the Test to 24 Different GOTV Treatments

We obtained our sample by identifying all GOTV field experiments published since 2000 in ten leading journals where the data was available online.[8] We augmented the sample by directly requesting data from authors. In principle we can apply our test to any GOTV effort. However, when the average effect of a get-out-the-vote effort is zero, we should not expect to find any interactive effect unless the treatment demobilizes some subset of individuals. For this reason, we restrict our analysis to available experiments with positive and statistically significant average treatment effects.[9] Even within a particular study, we only analyze those experimental treatments which exhibit a statistically significant effect on the average level of voter turnout. Table 2 presents a summary of the published studies utilized in this study. For each study, the table reports the delivery method of the experimental treatment, the electoral context, and the set of covariates available for the calculation of our propensity variables. See the Online Appendix for more details on each study and the nuanced aspects of our analyses in each setting.

[Table 2]

The regression results for each treatment are presented in Table 3. For each experiment, the table presents the coefficient on the *Treatment* variable, indicating the effect of the treatment for the average citizen in the sample. More importantly, the table presents the coefficient on *Treatment*Propensity*, indicating the extent to which the treatment effect changes as propensity increases. Because the *Propensity* variable is recoded so that the mean equals 0 and the standard deviation equals 1, we can interpret the interactive coefficient as the extent to which the treatment effect increases for every standard deviation increase in propensity. In Figure 3, we show nonparametric analyses, similar to that in Figure 2, for each of these 24 interventions. Overall, we see that GOTV interventions tend to exacerbate the participation gap. The interactive coefficient is positive in 16 out of 24 cases and positive and statistically significant for 8 of these cases. Alternatively, we find significant evidence for a reduction in the participation gap for only 2 out of 24 interventions.

[Table 3 and Figure 3]

How should we interpret the 14 cases where we find no statistically significant interactive coefficient? In many cases, our test is underpowered and unlikely to detect a real, substantively significant effect. For example, in the Stonybrook experiment (N06), we are unable to statistically reject interactive effects as low as -7.5 percentage points or as high as 4.7 percentage points – in other words, the interactive effect could be hugely positive or negative. However, in other cases where we have more statistical power, our lack of a statistically significant coefficient is a sign that even if there is an interactive effect, its substantive magnitude is likely small. For example, for the professional phone bank experiment (N07), we can statistically reject any interactive effect that is less than -0.6 percentage points or greater than 0.9 percentage points. For this reason, we urge readers to interpret the substantive size of the coefficients and standard errors in Table 3 without overly emphasizing statistical significance.

In the final row of Table 3 we conduct a pooled analysis by combining observations from all experiments in a single regression. In total, 319,251 individuals received one of these GOTV treatments and 848,521 were assigned to a control group, receiving no such treatment. With these pooled data, we assess the overall effect of these GOTV treatments on the participation gap. For nearly 1.2 million individuals, we regress voter turnout on a treatment dummy variable, the interaction of the treatment with each propensity score, study fixed effects, and the interaction of each study with the propensity score. The inclusion of study dummy variables and study-propensity interactions is necessary because the treatment is random within each study but not between studies.[10] Overall, these treatments exhibit a large positive effect on voter turnout for the average individual, 3.3 percentage points. However, this treatment effect is much stronger for high-propensity individuals. For every standard deviation increase in propensity, the treatment effect increases by 0.5 percentage points, on average. This suggests that GOTV treatments, on average, increase turnout by 4.3 percentage points for those whose propensity score is 2 standard deviations above the mean. However, these treatment effects are much weaker, only 2.3 percentage points, for low-propensity citizens. On average, GOTV mobilization effects are more than 85 percent greater for the highest propensity individuals compared to the lowest. As a result, the typical treatment exacerbates the participation gap, despite the fact that GOTV interventions are often designed to reduce this gap. In the Online Appendix, we describe additional tests which ensure that our results are not driven by “deadweight” or other issues with the quality of data available on voter files.

What is the mechanism behind this effect?

Readers surprised by our results will naturally ask why these experiments tend to exacerbate the participation gap? Determining the mechanism is difficult to answer in any setting (Bullock, Green, and Ha 2010) but deserves attention nonetheless. Here, we provide several hypotheses and provide evidence that while high-propensity individuals are easier to contact, even within the contacted set, the

relationship between propensity to vote and response to treatment persists. First, the correlation of voting propensity with education and political knowledge might result in high-propensity citizens better understanding the treatment. Also, low-propensity citizens may have higher costs involved with voting so a treatment that reminds voters of the benefits of voting may have to be more powerful to stimulate low-propensity voters than high-propensity voters. Moreover, psychological, social, and economic differences between high and low-propensity citizens may explain why some people are simply more likely to comply with any policy or experimental intervention. Previous research shows that higher SES individuals are more likely to respond to many interventions including electoral reform (Berinsky 2005), public health campaigns (Pickett, Luo, and Lauderdale 2005), medical screenings (Wee et al. 2012), public housing (Blundell, Fry, and Walker 1988), and Medicaid (Aizer 2003) even when such interventions are specifically designed to benefit low SES individuals.

A related potential mechanism lies in citizens' differential probabilities of being successfully contacted. High-propensity individuals may be more likely to read their mail, answer their phone, or talk to a canvasser. With our data, we explicitly test whether high-propensity citizens are easier to contact via phone or door-to-door canvassing. For many studies, including direct mail studies, the researcher is unable to know who actually received the treatment, but for phone and canvassing studies, the researcher can record which individuals or households were contacted. The results of these tests for all available studies are available in the Online Appendix. On average, high-propensity individuals are much easier to contact via door-to-door canvassing or phone calls. For example, in Gerber and Green's (2000a) New Haven study, a standard deviation increase in propensity corresponds to an extra 5 percentage point chance of canvassing contact and an extra 12 percentage point chance of phone contact. These results suggest that differential contact rates may explain much of the variation in intention-to-treat effects between high and low-propensity voters. If subsequent interventions hope to improve the participation and representation of low propensity voters, they must first overcome the challenge of reaching them in the first place.

While differential contact rates are one primary reason that low-propensity citizens are more difficult to mobilize, they are not the only reason. Nickerson (2008) conducted door-to-door canvassing experiments with placebo treatments, allowing us to run separate analyses where we confine our observations only to individuals or households that were contacted by the experimenters. Because of the placebo treatments, we know which individuals in this control group would have been contacted by the canvassers had they been in the treatment group and we know that they are comparable in expectation to those who were contacted in the treatment group. When we apply our test to only the sample of individuals or households who answered the door, we still estimate large, positive interactive coefficients. Even among the sample of individuals or households willing to answer the door, high propensity citizens were more responsive to the get-out-the-vote message. More details on these tests are provided in the Online Appendix. While contact is a significant barrier to mobilizing low-propensity citizens, further barriers remain. Even conditional on receiving the message, low propensity citizens are less likely to respond to it.

Further information regarding the mechanisms behind our overall findings can be garnered by examining variation in our interactive effects across different interventions. In the Online Appendix, we explore several sources of this variation. Consistent with Arceneaux and Nickerson (2009), we find that the exacerbating effect of GOTV treatment is greatest in low-salience elections. However, even in high-salience elections, these treatments, on average, increase disparities in participation. We also find that the most effective treatments in terms of increasing overall participation also exhibit the greatest exacerbating effects. Finally, we discuss the peculiar phenomenon that among the studies we examine, the few experimental settings in which the participation gap was reduced involved largely African-American samples.

In one sense, why GOTV interventions tend to exacerbate the participation gap is of secondary importance. Whether the mechanism involves knowledge, costs, psychology, contact, or something else, the positive and normative implications of our study are the same. However, knowing that low-

propensity citizens are harder to contact provides one promising avenue for future researchers to design interventions that may mobilize them. Reaching the population of interest would significantly mitigate (but not remove entirely) the exacerbating phenomenon that we identify. We hope that the continued application of our test on future experiments will allow us to understand the reasons why and the conditions under which campaign interventions exacerbate or reduce the participation gap.

Conclusion

In analyzing 24 field experiments, we find that two-thirds of GOTV experiments mobilized high-propensity voters to a greater degree than low-propensity voters – thereby exacerbating the participation gap. Moreover, this exacerbating effect is statistically significant in 8 cases. Our pooled analysis demonstrates an average exacerbating effect that is substantively large and statistically significant. On average, GOTV methods developed by political scientists, many of which have been subsequently adopted by political campaigns (Gerber et al. 2011; Green and Gerber 2008; Issenberg 2010, 2012), appear to exacerbate the disparities between voters and the voting-eligible population. Because turnout is politically consequential and because our *propensity* variable captures meaningful political differences between individuals and groups, this exacerbating effect should be of concern to both political scientists and practitioners.

These results pose and clarify a challenge to scholars and practitioners interested in political participation. Current mobilization methods employed by researchers and political campaigns appear to be mainly effective in bringing high propensity citizens to the polls. The search for a reliable mobilization method for bringing under-represented, low-propensity citizens to the polls continues, largely without resolution, and these results underscore the need for an increased focus on this goal. The method we present for assessing differential treatment effects provides a metric for future experimenters interested in the effects of GOTV interventions on the composition of the electorate. We hope that this tool will aid in the development and assessment of new GOTV methods that are

effective in mobilizing low-propensity citizens and reducing the demographic gaps between the electorate and the greater, eligible population.

Even for those uninterested in the participation gap *per se*, our findings hold practical implications for campaigns and scholars of voter mobilization. Some individuals, particularly those with low propensities, are harder to mobilize than others. As a result, the combinatorial effects of multiple GOTV efforts are called into question. Practitioners may wrongly assume that the effects of one mobilization will combine additively with the effects of others. However, once a campaign has hit a ceiling among high-propensity voters, it may be unable to further increase participation using traditional methods.

The findings of this study also raise an ethical concern for experimenters and practitioners because experimental interventions and mobilization efforts are often conducted with the assumption that raising average participation levels can only be good for democracy. However, the evidence in this article – that voter mobilization tends to exacerbate existing inequalities in the electorate – necessitates a more nuanced perspective. Despite their unquestionable accomplishments in raising mean levels of participation, current GOTV efforts are not the solution to persistent inequalities in the political process. On the contrary, these efforts may contribute to the problem by making the electorate more polarized and less representative of the greater population.

Acknowledgements

We thank Kevin Arceneaux, Steve Ansolabehere, Peter Aronow, Catherine Choi, Lisa Garcia Bedolla, Don Green, Andy Hall, Ben Lauderdale, Melissa Michelson, Joel Middleton, David Nickerson, Todd Rogers, Jim Snyder, Aaron Strauss, and all authors who made their data available to us.

Endnotes

[1] Our empirical strategy only allows us to assess the effect of randomized GOTV interventions, which have been primarily conducted by academic researchers often in conjunction with interest groups or campaigns. We cannot explicitly test for the effects of other GOTV interventions which were not randomized. Nonetheless, to the extent that these GOTV experiments are similar to other mobilization tactics, our results inform us about the effects of the broader voter mobilization enterprise outside of academic research.

[2] We use OLS, a linear probability model, for this step but the results are virtually unchanged if we use Logit or Probit instead. Advantages of the linear probability model in the setting as compared to other models include computational speed, fewer assumption, and the avoidance of the “incidental parameters problem” (Angrist and Pischke 2009).

[3] Future scholars who implement this test should be sensitive to these issues, particularly if the sample size is small.

[4] Given the natural variation in voter turnout within demographic subgroups, this r-squared statistic fits well within the bounds of r-squared statistics found in other published models of voter turnout. To obtain a better sense of the extent to which our propensity model captures meaningful differences within the electorate, we can examine the range and distribution of the propensity variable before rescaling. The top panel of Figure 2 shows that our propensity variable identifies some subgroups with probabilities of voting as low as .1 or as high as .6. Results for the other experiments, shown in the Online Appendix, are similar.

[5] Standard errors are clustered by household (or block) where applicable and heteroskedasticity-robust in all cases. See the Online Appendix for more details.

[6] Ansolabehere, Stephen, 2011-4-13, “CCES Common Content, 2008”,

<http://hdl.handle.net/1902.1/14003 V4>.

[7] Jackman, Simon and Lynn Vavreck. 2010. Cooperative Campaign Analysis Project, Release 2.1. YouGov, Inc., Palo Alto, CA.

[8] Our list of journals was comprised of *American Journal of Political Science*, *American Political Science Review*, *American Politics Research*, *British Journal of Political Science*, *Electoral Studies*, *Journal of Politics*, *Political Analysis*, *Political Behavior*, *Political Research Quarterly*, and *Quarterly Journal of Political Science*.

[9] While the demobilization of a subset of GOTV subjects is theoretically possible (e.g. Mann 2010), we find no systematic evidence of this. We checked for interactive effects in the experiments reported in Nickerson (2007a), Vavreck (2007), Green and Vavreck (2008), Gerber, Karlan, and Bergan (2010), and Shaw et al. (2012), all of which had close to zero average effect, and we found no interactive effects.

[10] Suppose 50% of subjects are assigned to treatment in one experiment and only 10% of subjects are assigned to treatment in another. The treated and control individuals would not be comparable between studies, but they are comparable within studies. To prevent the different treatment rates from influencing our estimates, we include study fixed effects along with study-propensity interactions. This ensures that only the within-study variation (that which is randomly assigned) contributes to our estimates.

References

- Aizer, Anna. 2003. Low Take-up in Medicaid: Does Outreach Matter and for Whom? *American Economic Review Papers and Proceedings* 3(2):238-241.
- Alvarez, R. Michael, Asa Hopkins, and Betsy Sinclair. 2010. Mobilizing Pasadena Democrats: Measuring the Effects of Partisan Campaign Contacts. *Journal of Politics* 72(1):31-44.
- American Political Science Association Task Force. 2004. American Democracy in an Age of Rising Inequality. *Perspectives on Politics* 2(4):651-666.
- Angrist, Joshua and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press.
- Anzia, Sarah. 2011. Election Timing and the Electoral Influence of Interest Groups. *Journal of Politics* 73(2):412-427.
- Arceneaux, Kevin and David Nickerson. 2009. Who is Mobilized to Vote? A Re-Analysis of 11 Field Experiments. *American Journal of Political Science* 53(1):1-16.
- Avery, Michael. 1989. *The Demobilization of American Voters*. Greenwood Press.
- Bartels, Larry. 2008. *Unequal Democracy*. Princeton University Press.
- Bartels, Larry. 2009. Economic Inequality and Political Representation. *The Unsustainable American State*, eds. Lawrence Jacobs and Desmond King. Oxford University Press.
- Blundell, Richard, Vanessa Fry, and Ian Walker. 1988. Modelling the Take-up of Means-tested Benefits: The Case of Housing Benefits in the United Kingdom. *The Economic Journal* 98:58-74.
- Berinsky, Adam. 2005. The Perverse Consequences of Electoral Reform in the United States. *American Politics Research* 33(4):471-491.
- Bullock, John, Donald Green, and Shang Ha. 2010. Yes, But What's the Mechanism? (Don't Expect an Easy Answer). *Journal of Personality and Social Psychology* 98(4):550-558.
- Citrin, Jack, Eric Schickler, and John Sides. 2003. What if Everyone Voted? Simulating the Impact of Increased Turnout in Senate Elections. *American Journal of Political Science* 47(1):75-90.

- Dahl, Robert. 2006. *On Political Equality*. New Haven: Yale University Press.
- Dale, Allison and Aaron Strauss. 2009. Don't Forget to Vote: Text Message Reminders as a Mobilization Tool. *American Journal of Political Science* 53(4):787-804.
- Fowler, Anthony. 2013. Electoral and Policy Consequences of Voter Turnout: Evidence from Compulsory Voting in Australia. *Quarterly Journal of Political Science* 8(2):159-182.
- Gabler, Nicole, Naihua Duan, Diana Liao, Joann Elmore, Theodore Ganiats, and Richard Kravits. 2009. Dealing with Heterogeneity of Treatment Effects: Is the Literature Up to the Challenge? *Trials* 10:43.
- Garcia Bedolla, Lisa and Melissa R. Michelson. 2012. *Mobilizing Inclusion: Transforming the Electorate through Get-out-the-vote Campaigns*. New Haven: Yale University Press.
- Gerber, Alan and Donald Green. 2000a. The Effects of Canvassing, Direct Mail, and Telephone Contact on Voter Turnout: A Field Experiment. *American Political Science Review* 94:653-663.
- Gerber, Alan and Donald Green. 2000b. The Effect of a Nonpartisan Get-Out-The-Vote Drive: An Experimental Study of Leafletting. *Journal of Politics* 62(3):846-57.
- Gerber, Alan, Donald Green, and David Nickerson. 2003. Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments. *Journal of Politics* 65(4):1083-1096.
- Gerber, Alan, Donald Green, and Christopher Larimer. 2008. Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *American Political Science Review* 102(1):33-48.
- Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions". *American Economic Journal: Applied Economics* 1 (2): pp. 35-52.
- Gerber, Alan, Donald Green, and Christopher Larimer. 2010. An Experiment Testing the Relative Effectiveness of Encouraging Voter Participation by Inducing Feelings of Pride or Shame. *Political Behavior* 32(3):409-422.

- Gerber, Alan, Gregory Huber, and Ebonya Washington. 2010. Party Affiliation, Partisanship, and Political Beliefs: A Field Experiment. *American Political Science Review* 104(4):720-744.
- Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. 2011. "How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment". *American Political Science Review* 105 (1): pp. 135-150.
- Gilens, Martin. 2005. Inequality and Democratic Responsiveness. *Public Opinion Quarterly* 69(5)778-796.
- Gilens, Martin. 2012. *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton University Press.
- Green, Donald P. and Alan S. Gerber. 2008. *Get out the Vote: How to increase voter turnout*. Brookings Press.
- Green, Donald and Holger Kern. 2012. Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees. *Public Opinion Quarterly*, forthcoming.
- Green, Donald and Melissa R. Michelson. 2009. "ACORN experiments in minority voter mobilization" in *The People Shall Rule*, Robert Fisher (ed). Nashville: Vanderbilt University Press.
- Green, Donald P. and Lynn Vavreck. 2008. Analysis of Cluster-Randomized Field Experiments: A Comparison of Alternative Estimation Approaches. *Political Analysis* 16: (2), pp. 138-152.
- Griffin, Brian and Brian Newman. Are voters better represented. *The Journal of Politics* 67 (4), pp. 1206-1227.
- Hajnal, Zoltan and Jessica Trounstine. 2005. "Where turnout matters: the consequences of uneven turnout in city politics". *The Journal of Politics* 67(2):515-536.
- Hansford, Thomas and Brad Gomez. 2010. Estimating the Electoral Effects of Voter Turnout. *American Political Science Review* 104(2):268-288.
- Issenberg, Sasha. 2010. "Nudge the Vote." *The New York Times Magazine*, October 29, 2010.

- Issenberg, Sasha. 2012. *The Victory Lab: The Secret of Winning Campaigns*. New York: Crown.
- Key, V. O. 1949. *Southern Politics in State and Nation*. Multiple presses.
- Lijphart, Arend. 1997. Unequal Participation: Democracy's Unresolved Dilemma. *American Political Science Review* 91(1):1-14.
- Mann, Christopher. 2010. Is There Backlash to Social Pressure? A Large-scale Field Experiment on Voter Mobilization. *Political Behavior* 32(3):387-407.
- Michelson, Melissa R., Lisa García Bedolla and Donald P. Green. 2008. "Mobilizing the Registered: Tactics for Improving Voter Turnout." In Morgan E. Felchner, ed., *Voting in America*, (Praeger), pp. 198-211.
- Middleton, Joel and Donald Green. 2008. Do Community-Based Mobilization Campaigns Work Even in Battleground States? Evaluating the Effectiveness of MoveOn's 2004 Outreach Campaign. *Quarterly Journal of Political Science* 3(1):63-82.
- Nickerson, David. 2006. Volunteer Phone Calls Can Increase Turnout: Evidence from Eight Field Experiments. *American Politics Research* 34(3):271-292.
- Nickerson, David. 2007a. Does Email Boost Turnout? *Quarterly Journal of Political Science* 2: 369-379.
- Nickerson, David. 2007b. Quality is Job One: Professional and Volunteer Voter Mobilization Calls. *American Journal of Political Science* 51(2):269-282.
- Nickerson, David. 2008. Is Voting Contagious? Evidence from Two Field Experiments. *American Political Science Review* 102(1):49-57.
- Nickerson, David and Todd Rogers. 2010. Do You Have a Voting Plan? Implementation Intentions, Voter Turnout, and Organic Plan Making. *Psychological Science* 21(2)194-199.
- Pickett, Kate, Ye Luo, and Diane Lauderdale. 2005. Widening Social Inequalities in Risk for Sudden Infant Death Syndrome. *American Journal of Public Health* 95(11):1976-1981.

- Pocock, Stuart, Susan Assmann, Laura Enos, and Linda Kasten. 2002. Subgroup Analysis, Covariate Adjustment, and Baseline Comparisons in Clinical Trial Reporting: Current Practice and Problems. *Statistics in Medicine* 23:2509-2525.
- Putnam, Robert D. 2000. *Bowling Alone*. New York: Simon and Schuster.
- Schattschneider, E. E. 1960. *The Semisovereign People: A Realist's View of Democracy in America*. Holt, Rinehart, and Winston.
- Schlozman, Kay Lehman, Sidney Verba, and Henry E. Brady. 2012. *The Unevenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy*. Princeton University Press.
- Shaw, Daron, Donald Green, James Gimpel, and Alan Gerber. 2012. Do Robotic Calls from Credible Sources Influence Voter Turnout or Vote Choice? Evidence from a Randomized Field Experiment. *Journal of Political Marketing* 11(4):231-245.
- Vavreck, Lynn. 2007. "The Dangers of Self-Reports of Political Behavior", *Quarterly Journal of Political Science* 2:287.
- Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality*. Cambridge: Harvard University Press.
- Wee, Liang En, Gerald Choon Huat Koh, Wei Xin Yeo, Run Ting Chin, Jolene Wong, and Braden Seow. 2012. Screening for Cardiovascular Disease Risk Factors in an Urban Low-income Setting at Baseline and Post Intervention: A Prospective Intervention Study. *European Journal of Preventative Cardiology*, forthcoming.
- Wolfinger, Raymond E. and Steven J. Rosenstone. 1980. *Who Votes?* New Haven: Yale University Press.

Author Bios

Ryan D. Enos is Assistant Professor in the Department of Government at Harvard University, Cambridge, MA 02138.

Anthony Fowler is Assistant Professor in the Harris School of Public Policy Studies at the University of Chicago, Chicago, IL 60637.

Lynn Vavreck is Associate Professor of Political Science and Communication at the University of California at Los Angeles, Los Angeles, CA 90065.

Table 1. Does the propensity variable capture meaningful political characteristics?

	CCES	CCAP
Family Income	6117(305)**	6022(399)**
College	.065(.003)**	.059(.003)**
Church Attendance	.052(.003)**	.040(.005)**
Married	.097(.004)**	.041(.006)**
Employed or Retired	.086(.004)**	.068(.006)**
Bush Approval	.027(.003)**	.026(.004)**
Allow Abortion	.013(.004)**	.029(.004)**
Withdraw Troops from Iraq	-.033(.004)**	-.031(.006)**
5 Pt. Ideology (0 = Very Lib., 1 = Very Con.)	.013(.002)**	.010(.003)**
Extreme Ideology	.037(.003)**	.015(.005)**
7 Pt. Party ID (0 = Strong Dem., 1 = Strong Rep.)	.009(.003)**	.007(.004)*
Strong Party ID	.091(.004)**	.062(.005)**
Cut Domestic Spending	.014(.004)**	
Support Affirmative Action	-.018(.004)**	
Raise Minimum Wage	-.038(.003)**	
Ban Gay Marriage	.034(.004)**	
Support Federal Housing Assistance	-.020(.004)**	
Support Civil Unions		.006(.006)
Support Universal Healthcare		-.007(.006)
Raise Taxes on Wealthy Families		-.037(.005)**
Citizenship for Illegal Immigrants		.000(.006)
Sample Size	25,481	16,518

*Robust standard errors in parentheses; ** significant at 1%, * significant at 5%*

This analysis employs survey data from the 2008 CCES and 2008 CCAP studies to assess whether our propensity variable is meaningfully capturing politically relevant variables. The samples include verified registered voters who were successfully matched to statewide voter records. For both survey samples, we generate our propensity variable using validated turnout in the 2008 presidential election as our dependent variable, and using only variables available on voter files (age, race, gender, household size, state, and validated turnout in previous elections) as our independent variables. Then, we test whether this propensity variable is correlated with political characteristics and attitudes by regressing those variables on our propensity variable. Each cell in the table represents a separate regression of a dependent variable on the propensity variable. The first cell entry of 6,117 indicates that a single standard deviation increase in propensity corresponds with an increase in family income of \$6,117. Aside from income, ideology, and party ID, all other variables are coded as dummy variables. The propensity variable is positively related to income, education, church attendance, marriage, employment, Bush approval, conservative ideology, Republican party identification, having an extreme ideology, having a strong party identification, and a desire to ban gay marriage. Propensity is negatively correlated with a desire to withdraw troops from Iraq, support for affirmative action, a desire to raise minimum wage, support for federal housing assistance, and support for raising taxes on wealthy families. This analysis shows that when GOTV treatments increase the prevalence of high-propensity citizens in the electorate, they are meaningfully changing the demographic and attitudinal composition of the electorate.

Table 2. Summary of Studies and Available Data

Study	Method	Context	Available Covariates									
			Age	Race	Gender	Household Size	Geography	Vote History	Party Registration	Registration Year	Survey Responses	
Gerber and Green 2000a (GG00)	Multiple	1998 General - New Haven, CT	X					X	X	X		
Gerber, Green, Nickerson 2003 (GGN03)	Door	2001 Local - Multiple Cities	X	X	X				X	X		
Nickerson 2006 (N06)	Phone	2000 and 2001 - Multiple Cities	X		X			X	X	X	X	
Nickerson 2007b (N07)	Phone	2002 General - Multiple Cities	X			X		X	X	X		
Gerber, Green, Larimer 2008 (GGL08)	Mail	2006 Primary - Michigan	X		X	X			X			
Middleton and Green 2008 (MG08)	Door	2004 General - Multiple States	X		X			X	X			
Nickerson 2008 (N08)	Door	2002 Primary - Multiple Cities	X		X			X	X	X		
Dale and Strauss 2009 (DS09)	Text	2006 General - Multiple States	X	X	X			X	X		X	
Gerber, Green, Larimer 2010 (GGL10)	Mail	2007 Local - Michigan							X			
Gerber, Huber, Washington 2010 (GHW10)	Mail	2008 Primary - Connecticut	X		X	X			X		X	X
Nickerson and Rogers 2010 (NR10)	Phone	2008 Primary - Pennsylvania	X	X	X	X	X	X				

The table summarizes the available data employed for our study. Our data is drawn from 11 previously published studies, numerous different methods of delivery, and numerous types of electoral settings. To construct our propensity variable, we employ any pre-treatment variables that can help us to predict voter turnout: age, race, gender, household size, geography, vote history, party registration, registration year, and in one case survey responses.

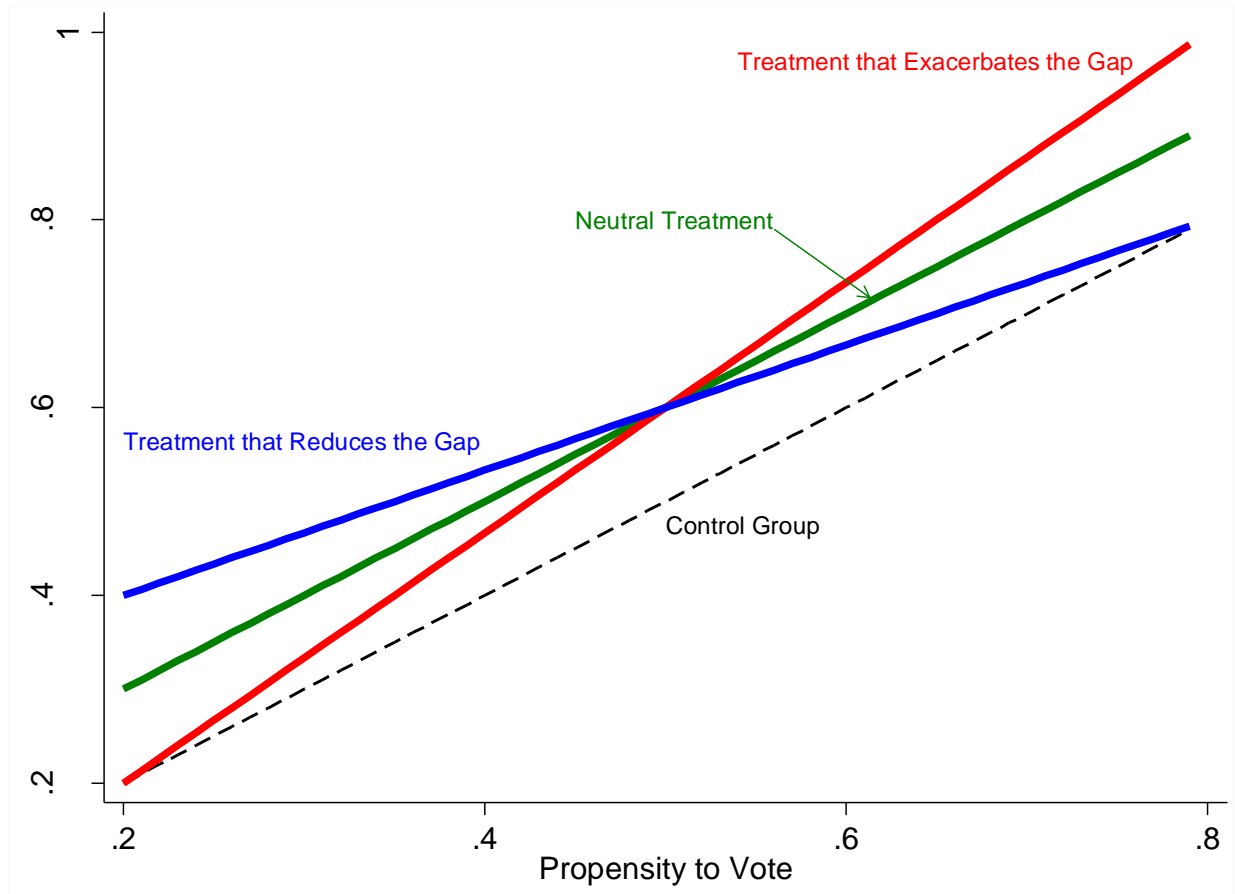
Table 3. Summary of Results

Intervention	Treatment	Treatment*Propensity	N-treated	N-control	Study
Mail	.016(.008)*	.002(.007)	7,679	11,665	GG00
Door	.040(.011)**	-.006(.009)	2,877	11,665	
Mail+Door	.037(.012)**	.004(.010)	1,853	11,665	
Phone+Mail+Door	.031(.015)*	.026(.013)*	1,207	11,665	
Bridgeport	.049(.020)*	.052(.025)*	895	911	GGN03
Detroit	.027(.009)**	-.020(.006)**	2,472	2,482	
Minneapolis	.027(.013)*	.027(.010)**	1,409	1,418	
St. Paul	.035(.016)*	-.015(.011)	1,104	1,104	
Stonybrook	.071(.031)*	-.014(.031)	680	279	N06
Volunteer	.008(.004)*	-.004(.004)	26,565	27,221	N07
Professional	.016(.004)**	.001(.004)	27,496	27,221	
Prof.+Vol.	.015(.004)**	-.003(.004)	27,452	27,221	
Civic Duty	.018(.003)**	.002(.003)	38,218	191,243	GGL08
Hawthorne	.025(.003)**	.008(.003)**	38,204	191,243	
Self	.048(.003)**	.008(.003)**	38,218	191,243	
Neighbors	.080(.003)**	.016(.003)**	38,201	191,243	
MoveOn	.016(.004)**	-.010(.005)*	23,384	22,893	MG08
Minneapolis	.038(.016)*	.051(.017)**	876	1,748	N08
Text Message	.030(.010)**	-.017(.010)	4,007	4,046	DS09
Civic Duty	.017(.008)*	.017(.009)	3,238	353,341	GGL10
Shame	.064(.006)**	.029(.007)**	6,325	353,341	
Pride	.041(.006)**	.005(.006)	6,307	353,341	
Party Reg.	.034(.008)**	.011(.010)	1,173	1,175	GHW10
Planning	.010(.004)**	.000(.003)	19,411	228,995	NR10
Pooled	.033(.001)**	.005(.001)**	319,251	848,521	

Robust standard errors in parentheses; ** significant at 1%, * significant at 5%

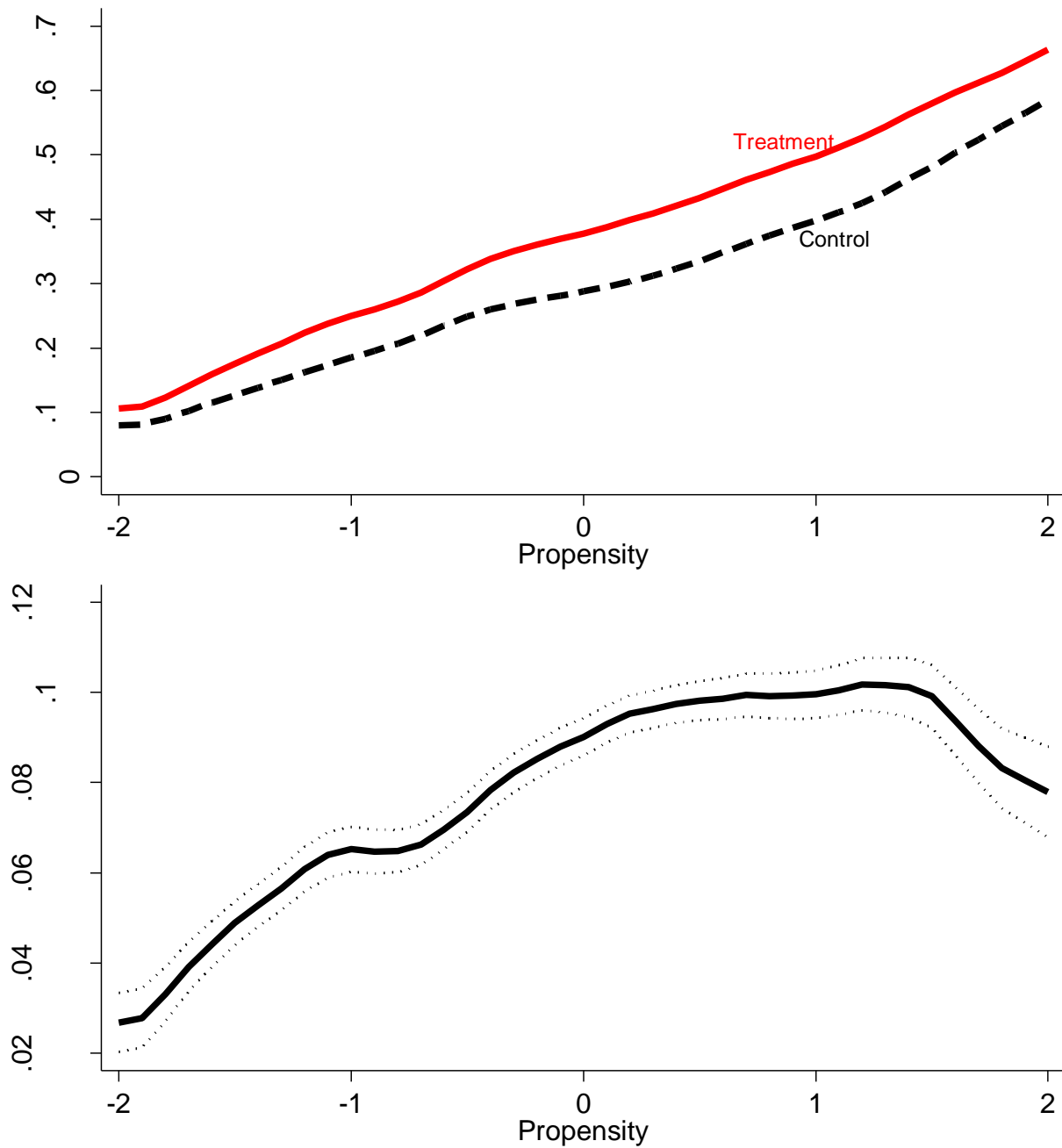
The table presents our regression results for 24 different experimental treatments. Standard errors are clustered by household (or block) where applicable and heteroskedasticity-robust in all other cases. We only include those treatments which demonstrate a statistically significant effect on the mean level of voter turnout. The Treatment coefficient indicates the treatment effect for the average citizen in the sample. The Treatment*Propensity coefficient indicates the extent to which the treatment effect changes as Propensity increases. Because the propensity variable is recoded so that the mean equals zero and the standard deviation equals 1, we can interpret the coefficient as the extent to which the treatment effect increases as propensity increases by 1 standard deviation. In 8 cases, we see a statistically-significant, positive coefficient, indicating that the treatment exacerbated in the participation gap. In only 2 cases do we see statistically-significant evidence that the participation gap was reduced. The final row presents a pooled analysis, showing the overall effect of GOTV experiments on the participation gap.

Figure 1. Three Hypothetical Experiments with the Same Average Effect



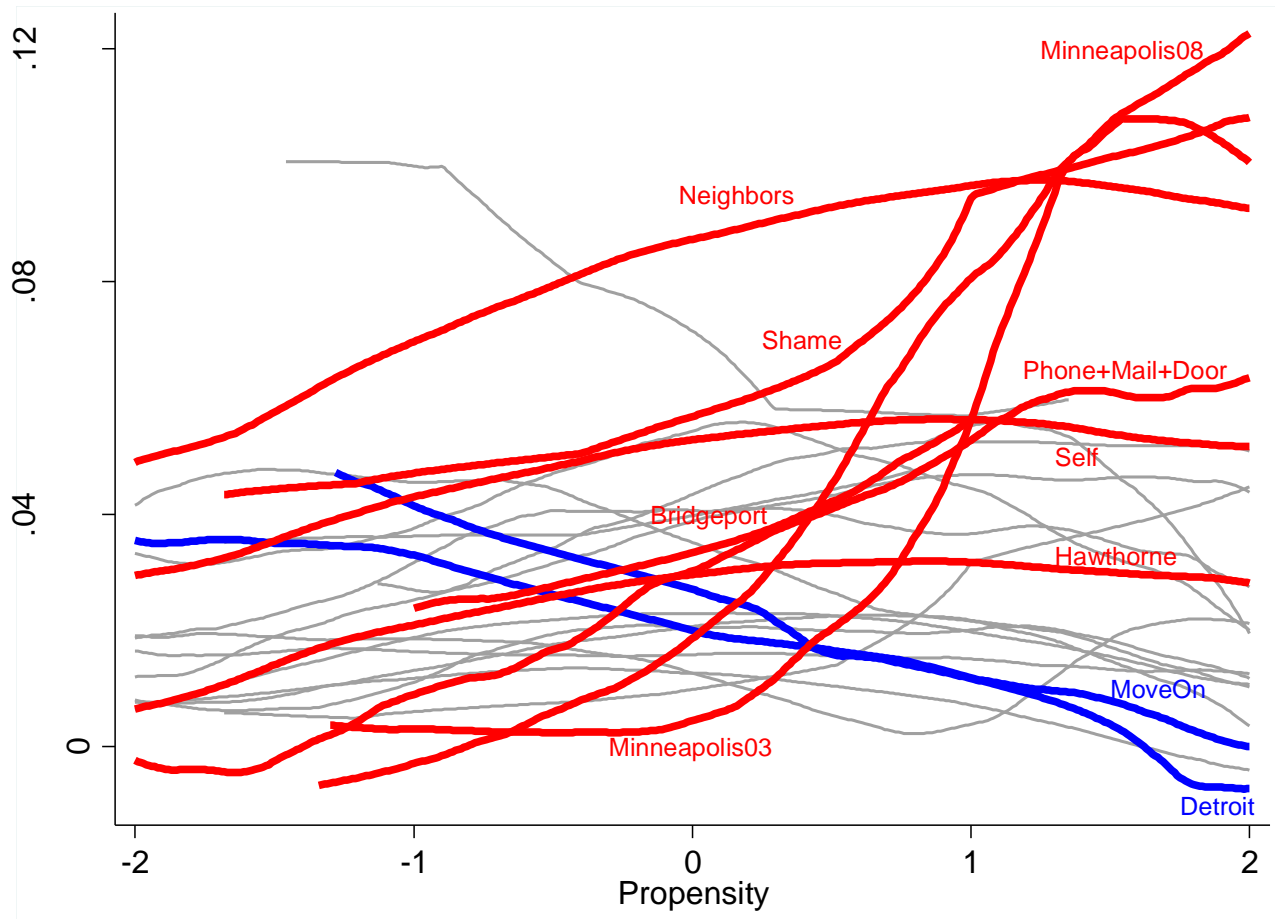
The figure presents three hypothetical experimental treatments. Each one significantly boosts the average level of turnout, but they have starkly different implications for the participation gap. The blue line reduces the participation gap by primarily mobilizing low-propensity citizens. The green line mobilizes citizens equally at all propensity levels and is therefore neutral in regard to the participation gap. The red line actually exacerbates the participation gap by primarily mobilizing high-propensity citizens, thereby making the electorate less representative of the voting-eligible population.

Figure 2. Variation in “Neighbors” Treatment Effect



The figure shows variation in the “neighbors” treatment across different propensity levels. The top panel presents kernel regressions of voter turnout on propensity separately for the treatment and control group. The bottom panel presents the difference between these two kernel regressions, presenting the conditional average treatment effect at each propensity level. The dotted lines indicate standard errors. We see that the “neighbors” treatment is much larger for high-propensity citizens relative to low-propensity citizens, thereby exacerbating the participation gap.

Figure 3. Variation in 24 GOTV Treatment Effects



The figure presents the conditional average treatment effect for all 24 experimental interventions across different levels of propensity. These plots mimic the bottom panel of Figure 2. The red curves indicate treatments which significantly exacerbate the participation gap, the blue curves indicate treatments which significantly reduce the participation gap, and the gray curves indicate treatments which had no statistically significant effect. The curves are cut off in order to avoid extrapolation. For example, if there are no observations in a particular setting with propensities below -1.5 , then we show no curve below that point.

Online Appendix

Increasing Inequality: The Effect of GOTV Mobilization on the Composition of the Electorate

Ryan D. Enos, Anthony Fowler, and Lynn Vavreck

This document provides additional information and results that are supplemental to the main paper. Additional details are provided about the data and statistical analyses. Also, additional results are provided regarding the mechanisms behind the main results and the variation of results across different experimental interventions. Please contact the authors with questions or comments.

Table of Contents

I. More Details on Each Analyzed Experiment	p. 2
II. More Details on the Pooled Analysis	p. 13
III. Are High-propensity Citizens Easier to Contact?	p. 15
IV. Does Contact Explain the Results Entirely?	p. 17
V. Can the Findings Be Explained by Data Limitations?	p. 19
VI. Variation Across Different Experimental Interventions	p. 21
VII. More Detailed Results	p. 26

I. More Details on Each Analyzed Experiment

Here we provide more details on each of the experiments that we analyzed. Because some of the studies included in our analysis contained unique features, readers may be interested to see how we incorporated those features into our analysis. We also report whether we can exactly replicate the original results reported by each published paper and compare our estimates of the additive coefficient from Table 3 (our estimated effect for respondents with an average propensity level) to the original studies' estimates of the average treatment effect. As we discuss in the text, these two estimates need not be the same. First, the average treatment effect is not the same thing as the effect for the average subject. Second, the treatment effect need not vary linearly across propensity levels (and we do not assume that it does). Rather, we employ our empirical specification to test whether the treatment effect increases or decreases, on average, as propensity increases. We also provide non-parametric analyses which allow us to assess the conditional average treatment effect as each level of propensity.

Gerber and Green (2000)

Gerber and Green (GG00) randomly assigned non-partisan phone calls, direct mail, and door-to-door canvassing across households in New Haven in the run up to the 1998 November general election. Households with 1 or 2 registered voters were randomly assigned to receive 0, 1, 2, or 3 mailings, 0 or 1 phone calls, and 0 or 1 door-to-door canvassing attempts. Some individuals received multiple treatments (e.g. 2 pieces of direct mail and a phone call). For the sake of simplicity, we collapse the direct mail conditions into 2 categories (no mail and at least one piece of mail). This leaves us with one control condition of 11,596 individuals and 7 different treatment conditions: Phone only ($n = 846$), Mail only (7,776), Door only (2,877), Phone+Mail (4,749),

Phone+Door (194), Mail+Door(1,853), and Phone+Mail+Door(1,207). These numbers match up very closely (but not exactly) with those in Table 2 in the original study.

For each of these 7 treatment groups, we test whether the treatment significantly increased voter turnout by comparing turnout treatment group to the control group. We focus only on intention-to-treat effects and do not incorporate data on successful contacts in our main analysis. Simple regressions of turnout on the treatment with controls for prior turnout, age, and party (which are not necessary but increase precision) reveal statistically significant treatment effects for only 4 of the treatment conditions: Mail, Door, Mail+Door, and Phone+Mail+Door. Some of the other treatment combinations mostly likely did exhibit a positive effect, but the sample size of the treatment group was too small to derive precise estimates. For example, we estimate that the Phone+Door intervention increased turnout by 5.6 percentage points, but the small sample size for the treatment group means that the estimate is not precise enough to include in our sample.

While Gerber and Green do not categorize their treatment groups in the same way that we do here, our analysis produces results that are very consistent with the results reported by the original study. We find no detectable effect of phone calls (with a slightly negative point estimate), a small positive effect of direct mail (about 1 percentage point), and a larger positive effect of door-to-door canvassing (3 to 4 percentage points). We also see evidence suggesting that these effects are not additive. For example, The Mail+Door treatment appears to be no more effective than the Door treatment alone, even though the Mail treatment on its own exhibited a positive effect. The average treatment effects (estimated by us and in GG00) line up very closely with our estimates of the “Treatment” coefficients (the effect for citizens with average propensity) in Table 3. Because approximately half of GG00’s subjects lived in two-person households (as opposed to one-person households), we cluster the standard errors by household in our analysis of this data.

Gerber, Green, and Nickerson (2003)

Large-scale door-to-door canvassing experiments were conducting in 6 cities, Bridgeport, Columbus, Detroit, Minneapolis, and St. Paul, leading up to the 2001 local elections in those cities. Gerber, Green, and Nickerson (GGN03) show statistically significant treatment effects in Bridgeport, Detroit, and St. Paul. We also find a positive, statistically significant treatment effect in Minneapolis after adding controls and improving the efficiency of the test, so we include that experiment in our analysis as well. Re-analyzing their data, we construct our propensity variable using age, party registration, gender, race, and turnout in 1999, 2000, and a 2001 primary. Not all variables are available for each city. For example, racial data is only available in Raleigh, and a 2001 primary was not conducted in Raleigh.

In each city, households were selected for contact and one registered voter within each household was specifically selected for study. The households were grouped geographically into “walk lists.” In some cases, treatment was randomly assigned across all households in the study, and in other cases, the randomization was stratified within walk lists to improve balance. GGN03 estimate the treatment effect in each city with a regression that includes walk list fixed effects. These fixed effects improve efficiency but are not necessary for unbiased results because, even in the cases of stratifications, the probability of treatment was uniform across all walk lists. For our own regression results in Table 3, we do not include fixed effects for walk lists, but the results are nearly identical in either case.

With the replication data from GGN03, we are able to exactly replicate their results shown in Table 2 (of the original study). Moreover, the additive effects shown in our Table 3 line up fairly closely with the average treatment effects estimated in the original study. Because the study only includes 1 registered voter per household, the standard errors do not need to be clustered. Consistent with GGN03, we report heteroskedasticity-robust standard errors.

Nickerson (2006)

Nickerson analyzed 8 different phone-based get-out-the-vote experiments across 6 cities in either the 2000 presidential election or in 2001 local elections. Only one of those experiments – that in Stonybrook, NY in the run up to the 2000 presidential election – showed a statistically significant effect of the treatment. Therefore, we only include this experiment in our analysis. We replicate Nickerson’s estimate of the average treatment effect in Stonybrook (8.2 percentage points), and our estimate of the additive coefficient in Table 3 (7.1 percentage points) matches this estimate closely. This particular experiment does not present an ideal opportunity for our analysis for several reasons. First, the sample size is small (680 individuals in the treatment group and 279 in the control group) leading to high standard errors and imprecise estimates. Second, the sample consists of newly registered college students meaning that there is no vote history data available, and little variance in age. As a result of these factors, there is little variance in our estimates of vote propensity for this sample and the standard error on our estimate of the interactive coefficient in this sample is significantly greater than in any other experiment in our analysis. For these reasons, we should not be surprised by the “null” finding in this case and should not put significant weight on this particular experiment.

Nickerson (2007)

This study assesses the effect of get-out-the-vote phone calls to young registrants in selected cities in the 2002 general election. The sample was randomly divided into four equally sized treatment groups: a control group which received no calls, a group that was called by a volunteer phone bank, another group called by a professional phone bank, and a final group called by both phone banks. We treat the three different treatment groups as three separate interventions and

estimate their effects by comparing turnout in each treatment group to that in the control group. Nickerson also embedded several sub-experiments into the study that we ignore or average over. The timing of GOTV calls and the content of conversation were randomized for some subjects, but we focus on the average effects of these three interventions.

Nickerson does not explicitly test for the average intent-to-treat effect for each of these three interventions. However, our own analysis suggests that all three interventions positive, statistically-significant average effects, although consistent with the original study, the average effect is smallest for the “volunteer only” intervention.

Because the treatment probabilities were constant across sites, site fixed effects are not necessary for unbiased estimates but they could potentially improve efficiency. In Table 3, we report our results without site fixed effects, but the results are unchanged if they are included. Dummy variables for sites are included in our propensity regression, so this geographic information is implicitly already contained in the “propensity” variable.

Gerber, Green, and Larimer (2008)

The main text provides details on Gerber, Green, and Larimer’s “neighbors” experiment. As part of that study, the authors also conducted three additional experimental interventions, all delivered via direct mail. A “self” treatment contained everything in the “neighbors” treatment except for information about other individuals. A “Hawthorne” treatment removed vote history information but emphasized that “YOU ARE BEING STUDIED.” Finally, a “civic duty” treatment removed the social pressure and monitoring components and provided only the more traditional encouragement: “DO YOUR CIVIC DUTY AND VOTE!” The four treatment groups were approximately equal in size (about 38,200 individuals in each) while the treatment group was larger (191,243 individuals). We examine each of the four interventions separately. All treatments

exhibited a positive, statistically significant average treatment effect, so all interventions are included in our analysis.

We successfully replicated the original results reported by the authors and generated our propensity variable as described in the main text. Because the treatments occurred at the household level, we cluster our standard errors by household. The authors stratified their randomization by geographic clusters which were determined by mail routes. However, the probabilities of treatment were constant across clusters so simple differences-in-means yield unbiased estimates of the average treatment effects. Cluster fixed effects could potentially be added to our regressions in order to improve efficiency, but our results are nearly identical with our without these fixed effects.

Middleton and Green (2008)

This study is the only study in our analysis that is not an explicitly randomized trial. Middleton and Green take advantage of the quasi-random inability of of MoveOn to treat certain neighborhoods where door-to-door canvassing was planned. They only include in their analysis streets on the border of two neighborhoods where one received treatment and one quasi-randomly did not receive treatment. Therefore, the houses on one side of the street were treated and houses on the other side of the street were not. Even though the unit of analysis is an individual, the randomization effectively took place at the level of a street. With their replication data in hand, our analysis is strikingly similar to that of our other studies despite the fact that this treatment was not explicitly randomized by the researchers. We were able to replicate the authors' original results.

In constructing our propensity scores, we only include treatment observations (as always) and we use the exact same set of pre-treatment variables used by Middleton and Green as control variables. Because the randomization took place at the street level and because the probability of treatment may slightly differ across streets (although not significantly), we include street fixed effects

in our final analysis (M&G call this variable “block_id”). Because the relevant unit receiving treatment is a side of a street, we cluster our standard errors by each group of individuals living on one side of a street (M&G call this variable “block_id_new”). All of these steps are consistent with the original analysis conducted by the authors to estimate the average treatment effect. Our “treatment” coefficient of .016 almost perfectly matched the original study’s estimates of the average intent-to-treat effect.

Nickerson (2008)

This study employed door-to-door canvassing experiments in the run up to primary elections in 2002 in Denver and Minneapolis. Households were evenly divided into one of three treatment groups: a no-contact control group, a placebo treatment group where individuals who answered the door were reminded to recycle, and a get-out-the-vote treatment group. For our main analysis, we collapse the placebo and no-contact groups into a single control group. We find no statistically-significant average intent-to-treat effect in Denver but we do find such an effect in Minneapolis, so we confine our analysis to the Minneapolis study.

Nickerson does not report the average ITT effects, because they are less relevant for the specific questions being addressed in the original study. Instead, the original paper focuses on differences in turnout between households and individuals contacted in the GOTV and placebo treatment groups. We take advantage of this novel design to address an additional question about whether high propensity citizens are easier to mobilize conditional on being contacted in the first place. That analysis is described in a subsequent section in the Appendix.

Dale and Strauss (2009)

The original study examines the effect of get-out-the-vote text messages in the 2006 general election. Half of the subjects were randomly assigned to a control condition of no contact. The remaining half received one of four types of text messages. Among the treatment group, the researchers randomly varied whether they assigned a “civic duty” or a “close election” treatment and they also randomly varied whether they included a phone number for a hotline where subjects could receive information about their polling location, creating four equally sized treatment groups. At the end of the study, the authors successfully matched approximately 8,000 subjects to voter files, leaving one control group of about 4,000 individuals and 4 treatment groups of about 1,000 individuals. When we analyze the average effect of each treatment separately, we only find a statistically significant estimate in one case (close election message with no hotline information). However, the point estimates are similar for all four treatments and the small sample sizes limit our precision. For that reason, we pool all treatments together and estimate the average effect of any text message. Consistent with the original study, we estimate an effect of 3.0 percentage points, and that estimate is statistically distinguishable from zero ($p = .01$). Applying our test to this data, the additive coefficient is nearly identical to this estimate of the average treatment effect.

Gerber, Green, and Larimer (2010)

This study represents an extension of the previous study by the same authors discussed earlier. The authors conducted a direct mail experiment in the run up to local election in Michigan in 2007. They focused on households with only one or two individuals in each household where all individuals had voted in a recent election and had abstained from another recent election. They randomly assigned households into one of four groups: a control group that received no contact (353,341 individuals), a “civic duty” treatment group similar to that in GGL08 (3,238 individuals), a

“shame” treatment group similar to the “self” treatment from GGL08 but where subjects were only informed about the recent election where they abstained (6,325 individuals), and a “pride” treatment group also similar to the “self” treatment but where subjects were only informed about the recent election where they voted (6,307). We replicate positive, statistically-significant treatment effects for all three treatments and apply our test separately to each treatment. Because the treatment was assigned at the household level, our standard errors are clustered by household. The additive coefficients are nearly identical to the average treatment effects reported in the original study.

Gerber, Huber, and Washington (2010)

In the original study, the authors focused on registered voters in Connecticut who were not registered with a particular party and who reported in a survey that they were politically independent (leaners on the 7 point party identification questions are included in the experiment). While the original authors focus most of their analysis on these leaners or “latent partisans,” we also include pure independents in our analysis as they were included in the experiment. In the 2008 presidential primary in Connecticut, voters had to be registered with a party in order to vote in the election, so the subjects would have to register with a party in order to even be eligible to turn out. As such, voter turnout was very low among the control group (2.5%). Among the 2,348 individuals deemed appropriate for the experiment (most of whom lived in one-person households), the authors randomly assigned half of them into the treatment group where they received mail informing them that they would have to identify with a party in order to vote in the primary and providing a blank party affiliation form. As expected, the treatment appeared to increase turnout in the primary election (turnout in the treatment group was 5.9 percentage points). Applying our test to the data, our additive coefficient lines up very closely with the average treatment effect. The minor discrepancy between our estimate and the estimate in Table 3 of the original study is explained by

the fact that the treatment effect was smaller for pure independents whom we include in the analysis but are not included in the original analysis.

Nickerson and Rogers (2010)

This study presents the study of registered Democrats in Pennsylvania in the 2008 presidential primary. Individuals without phone number or on a “do not call” list were excluded for practical purposes. Also for practical purposes, household with more than 3 registered voters were excluded. Additionally, individuals who had voted in multiple recent primaries were dropped because they were deemed to be extremely likely to vote in the absence of any treatment. Subjects were randomly divided into a control group (no contact) or one of three treatment groups where their household received a phone call. In the first group, subjects received a traditional GOTV phone call reminding them about the election and priming their civic duty. Individuals in the second group received the standard GOTV treatment and were also asked whether they planned to vote. Individuals in the final group received these 2 treatments and were also asked three follow-up questions designed to facilitate plan-making. The standard GOTV and the self-prediction treatments had no statistically significant effect over the control condition, so these two treatments are removed from our analysis. We focus solely on the “planning” treatment.

The randomization was stratified by household size. Also, while this fact is not reported in the paper, the probability of treatment varied across household size (with a lower treatment probability for one-person households). Therefore, household fixed effects are necessary for unbiased estimates. For this reason, we include household fixed effects in both the estimation of propensity scores and the subsequent estimation of the interactive effect. As it turns out, results are nearly identical if we exclude the household fixed effects from the second estimation step because

the propensity variable already contains that information. Because the treatment occurs at the household level, we cluster our standard errors by household.

We are unable to exactly replicate the results reported in the original study, but our own estimates of the ITT effects (including fixed effects for household size) are very close to the reported estimates. Similarly, our additive coefficient in Table 3 and its corresponding standard error are nearly identical to the intent-to-treat results reported in the original paper. ‘

II. More Details on the Pooled Analysis

In conducting our pooled analysis, we aggregate the data from all experiments into one regression. The total sample size is 1,167,771 individuals clustered into 877,787 households. 319,251 individuals received some treatment while 848,521 were set aside in a control group. The totals do not match up with the sums of the treated and control columns in Table 3, because for some for several of the interventions, the control group is unchanged. To be clear, we do not “double count” control individuals in our pooled analysis. For example, the 11,665 control individuals from Gerber and Green (2000) are only included once, along with the individuals from the various treatment groups.

For each individual in this analysis, we have a binary indicator for voter turnout, a binary indicator for experimental treatment. All treatments are treated equally in the analysis in order to obtain average estimates across all experiments. However, we do not assume that treatment effects are homogeneous across setting. We also have the propensity scores calculated previously. These scores are not necessarily comparable across settings. For example, an individual with a score of 0.5 in the GG00 sample is not necessarily comparable to an individual with a score of 0.5 in the GGL08 sample. Similarly, the experimental settings vary across low and high salience elections. To account for these factors, we include dummy variables for each study and interactions of the propensity variable with these dummy variables.

A final complicating factor with the pooled analysis is that two of the experimental studies require additional conditional variables in order to obtain unbiased estimates. In the case of Middleton and Green (2008), we must also include block fixed effects and in the case of Nickerson and Rogers (2010) we must include household size fixed effects. These covariates are included in the pooled analysis but they make almost no difference. We include these factors by creating a unique study dummy variable for each block within the Middleton and Green study and for each

household size within the Nickerson and Rogers study. See the replication data files for the code required to execute this test.

III. Are High-Propensity Citizens Easier to Contact?

Using all experimental data where contact information is available, we test whether high-propensity citizens are easier to contact via door-to-door canvassing or phone calls, helping us to understand whether differential contact rates can partially explain the differential intent-to-treat effects that we find. Each row of Table A1 represents a separate regression, where we regress a dummy variable for household contact on our propensity variable. Only individuals in the treatment group are included in these analyses. The *Propensity* coefficients as the extent to which a single standard deviation increase in propensity corresponds to the probability of household contact. For example, in Gerber and Green's (2000) New Haven study, a standard deviation increase in propensity corresponds to an extra 5 percentage point chance of canvassing contact and an extra 12 percentage point chance of phone contact. On average, high propensity citizens are much easier to contact than low-propensity citizens, suggesting that differential contact rates are one important mechanism behind our empirical results.

Table A1. Are high-propensity citizens easier to contact?

	Study	Propensity	Constant
Door-to-Door Canvassing	GG00	.052(.007)**	.313(.007)**
	GGN03 - Bridgeport	.079(.018)**	.180(.015)**
	GGN03 - Columbus	.038(.013)**	.113(.011)**
	GGN03 - Detroit	.012(.006)*	.157(.007)**
	GGN03 - Minneapolis	.044(.009)**	.103(.009)**
	GGN03 - Raleigh	.019(.010)	.359(.012)**
	GGN03 - St. Paul	.051(.009)**	.127(.011)**
	N08 - Denver	-.005(.015)	.334(.016)**
	N08 - Minneapolis	.061(.021)**	.462(.024)**
Phone Calls	GG00	.121(.007)**	.344(.007)**
	N06 - Albany	-.005(.018)	.616(.017)**
	N06 - Boston	.038(.014)**	.554(.014)**
	N06 - Stonybrook	.020(.012)	.886(.012)**
	N07 - Professional	.031(.003)**	.386(.003)**
	N07 - Volunteer	.060(.003)**	.422(.003)**
	NR10	.030(.002)**	.248(.002)**

*Robust/ household-clustered standard errors in parentheses; ** significant at 1%, * significant at 5%*

For each experiment where contact information is available, we test whether contact rates are higher for high-propensity citizens by regressing contact on propensity for those individuals in the treatment group. The table shows that high-propensity citizens are much easier to contact via both door-to-door canvassing and phone calls. For example, in Gerber and Green's (2000) New Haven study, a standard deviation increase in propensity corresponds to an extra 5 percentage point chance of canvassing contact and an extra 12 percentage point chance of phone contact. These results suggest that differential contact rates explain much of the variation in intention-to-treat effects between high and low-propensity voters.

IV. Does Contact Explain the Findings Entirely?

We find that high-propensity individuals are much easier to contact through either phone calls or door-to-door canvassing, suggesting that a significant share of the heterogeneous treatment effects that we identify can be explained by differential contact rates. Here, we test whether differential contact rates can explain all of these patterns. We take advantage of the fact that one of the control groups in Nickerson (2008) received a placebo treatment, so we know which of those individuals would have received a treatment had they been in the GOTV condition. Table A2 shows the results of our empirical test for 4 different samples: all individuals in a household that was contacted in both Minneapolis and Denver and all individuals who were contacted themselves in both Minneapolis and Denver. In each case, the sample sizes are small and the statistical precision is limited, but the interactive coefficients are large and positive in each case. Even conditional on being contacted, the conditional average treatment effects are much greater for high-propensity citizens than they are for low-propensity individuals. This analysis suggests that differential contact cannot entirely explain the patterns uncovered in the paper. Low-propensity individuals are indeed harder to contact, but they are also harder to mobilize even after they have been contacted.

Table A2. Applying Our Test to Households and Individuals Who Received Treatment

	Contacted Households		Contacted Individuals	
	Minneapolis	Denver	Minneapolis	Denver
Treatment	.076 (.025)**	.067 (.029)*	.097 (.030)**	.079 (.034)*
Propensity	.218 (.025)**	.290 (.015)**	.223 (.026)**	.283 (.019)**
Treatment*Propensity	.041 (.027)	.025 (.021)	.039 (.034)	.030 (.027)
Constant	.146 (.016)**	.385 (.021)**	.137 (.019)**	.397 (.024)**
Observations	786	1,124	394	562
R-squared	.389	.380	.397	.361

*Standard errors are in parentheses, clustered by household in the case of contacted household and heteroskedasticity-robust in the case of contacted individuals; ** $p < .01$, * $p < .05$.*

V. Can the Findings be Explained by Data Limitations?

We may worry that our finding is a result of poor data quality from the public voter files typically employed in field experiments. Public voter records are often inaccurate and out of date. For example, an individual may have passed away or moved and could not possibly vote in an upcoming election, but a researcher would have no way of knowing this. We refer to these individuals as “deadweight,” because they shouldn’t be on the voter file at all, but the researcher hopelessly tries to mobilize them. Deadweight could be particularly troubling for our study if these ineligible individuals tend to be classified as low-propensity. What if the treatment effect is actually homogeneous across the eligible population, but many individuals categorized as low-propensity are actually deadweight, leading us to falsely conclude that GOTV interventions exacerbate the participation gap?¹

To address this concern, we perform a simple sensitivity analysis. We cannot entirely rule out concerns about deadweight, but we can determine how extensive the problem would have to be in order to drive our results. Pooling all experiments in our sample and including study fixed effects, we estimate an average treatment effect of 3.7 percentage points. Then, we break the sample into 20 subsamples according to each individual’s propensity score.² As expected, the conditional average treatment effect is larger for the higher propensity subsamples. The largest treatment effect we

¹ From the perspective of the participation gap, the reason that an individual cannot be mobilized is highly relevant. If a person is deceased, then they truly should not be in the sample. However, if the tendency to move and be unreachable by a political campaign is correlated with turnout as well as demographics and policy preferences, then the systematic tendency of GOTV treatments to miss these individuals will increase the participation gap.

² Our general results are robust to different numbers of subsamples.

observe for a subsample is 5.2 percentage points. If the treatment effects were truly homogeneous and our result were driven by deadweight in the voter file, then we could say that the true treatment effect for non-deadweight individuals in each sample would have to be at least 5.2 percentage points. Therefore, the minimum proportion of deadweight in our sample would have to be 30% ($1 - .037/.052$) in order for us to obtain the results that we do. Put another way, deadweight could only explain our results if deadweight individuals constitute at least 30% of these GOTV samples.

While voter records surely contain errors, this 30% figure is implausibly large.³ We demonstrate this by estimating the proportion of deadweight on a typical voter file. Campaigns and for-profit data vendors have a strong incentive to identify and remove deadweight from the file because targeting deadweight is costly. According to the data base of Catalist, a widely used political data services company based in Washington D.C., only 4% of the individuals on statewide voter files are classified as “deceased” or having a “bad address.” Focusing specifically on those individuals who voted in the most recent election, that number shrinks to 2.5%. Also, large-scale mail based surveys in Florida and Los Angeles County suggest that less than 10% of individuals on voter files are deadweight (Ansolabehere et al. 2010). Our sensitivity analysis indicates that data quality cannot reasonably be argued to explain our results, because the actual amount of bad data is much less than it would have to be in order to pose a threat to our inferences.

³ This is especially true in cases where the researchers selectively chose their sample to minimize deadweight. For example, Gerber, Green, and Larimer (2008) only include individuals who voted in 2004 in their sample, so the proportion of their sample that could have moved or passed away in the two year interim period is small.

VI. Variation across Different Experimental Interventions

Having seen that GOTV interventions tend to exacerbate the participation gap, on average, we would like to know whether this effect varies across different contexts, settings, or mobilization methods. Our strategy allows us to assess this variation. By conducting our test across many experiments, we hope to identify the types of interventions that are most effective (or least ineffective) in reducing the participation gap. Of course, even though this is the largest analysis of different types of interventions yet undertaken, the sample size prevents us from making strong claims about variation across different interventions. Nevertheless, we hope that our test here will guide future researchers in applying this test to identify the types of treatments that might effectively reduce the participation gap.

First, we test for variation across electoral salience. Arceneaux and Nickerson (2009) argue that high-propensity voters will be easier to mobilize in low salience elections and low-propensity voters will be easier to mobilize in high salience elections. To test this hypothesis, we compare empirical results across elections with different levels of voter turnout. Figure A1 plots the interactive coefficient from our regressions against the constant coefficients from the same regressions. The constant term indicates the predicted probability of turnout for the average subject in the control group. The interactive term indicates the extent to which the treatment exacerbated the participation gap. So, by looking for a relationship between these coefficients, we can see if interventions are more likely to increase the participation gap in low or high salience elections

The hypothesis of Arceneaux and Nickerson is largely confirmed: as electoral salience increases, the exacerbating effect of GOTV interventions decreases. However, looking at the graph, we would only expect an intervention to decrease the participation gap in elections with turnout of 50% or greater. Elections with 50% turnout are rare in the U.S. outside of presidential races. This

analysis suggests that GOTV interventions *can* actually reduce the participation gap in very high-salience elections, but these interventions have the opposite effect in most settings.

Next, we test for variation across the strength of treatment. We quantify severity using our treatment coefficient, the effect of the treatment for the average subject in the sample. Figure A2 plots the interactive coefficient against the additive coefficient from our analyses. We see that as the severity of the experimental treatment increases, the exacerbating effect also increases. One possible explanation is that the most effective treatments tend to have a psychological or social pressure component to them, as opposed to the more traditional GOTV messages. This psychological component which makes these interventions so effective may have a particularly concentrated effect among high-propensity citizens. As a result, the most effective interventions have the unexpected consequence of exacerbating the participation gap.

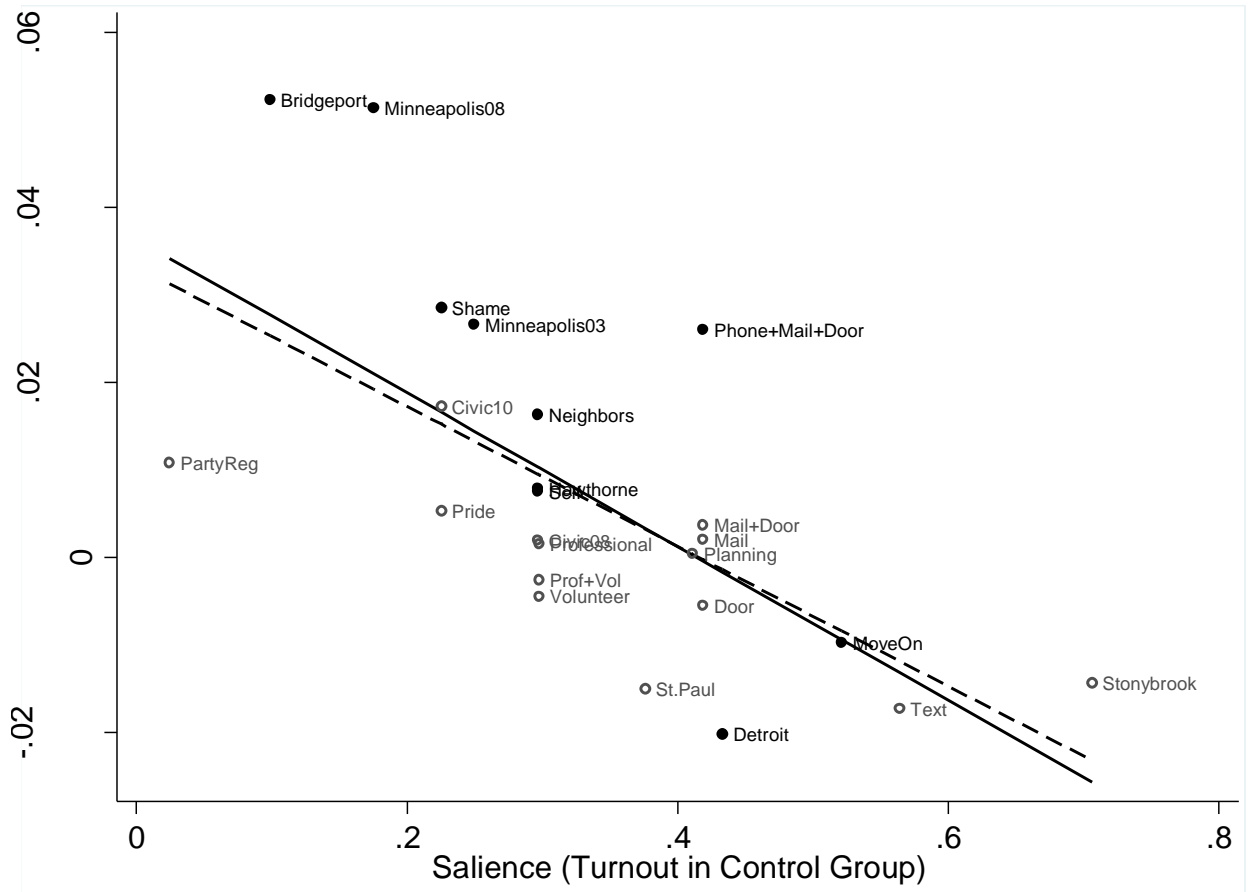
Only two interventions in our analysis demonstrate statistically significant evidence that the participation gap was reduced. What might explain the difference in these two cases? One intriguing similarity between the two experiments with negative interaction effects is that they both targeted citizens in communities with large African American populations. One explicitly targeted African Americans (Middleton and Green 2008) and the other was set in the largely African American city of Detroit⁴ (Gerber, Green, and Nickerson 2003).

We tested for the possibility that African Americans respond differently to GOTV experiments by examining field experiments for which both Blacks and non-Blacks are identified in the experimental population. Most public voter files do not identify the race of the voters. As such, there are only three studies in our sample that could be used for this purpose: Dale and Strauss (2009); the Gerber, Green, and Nickerson (2003) study in Raleigh; and Nickerson and Rogers

⁴ Detroit was 83% African American in 2010.

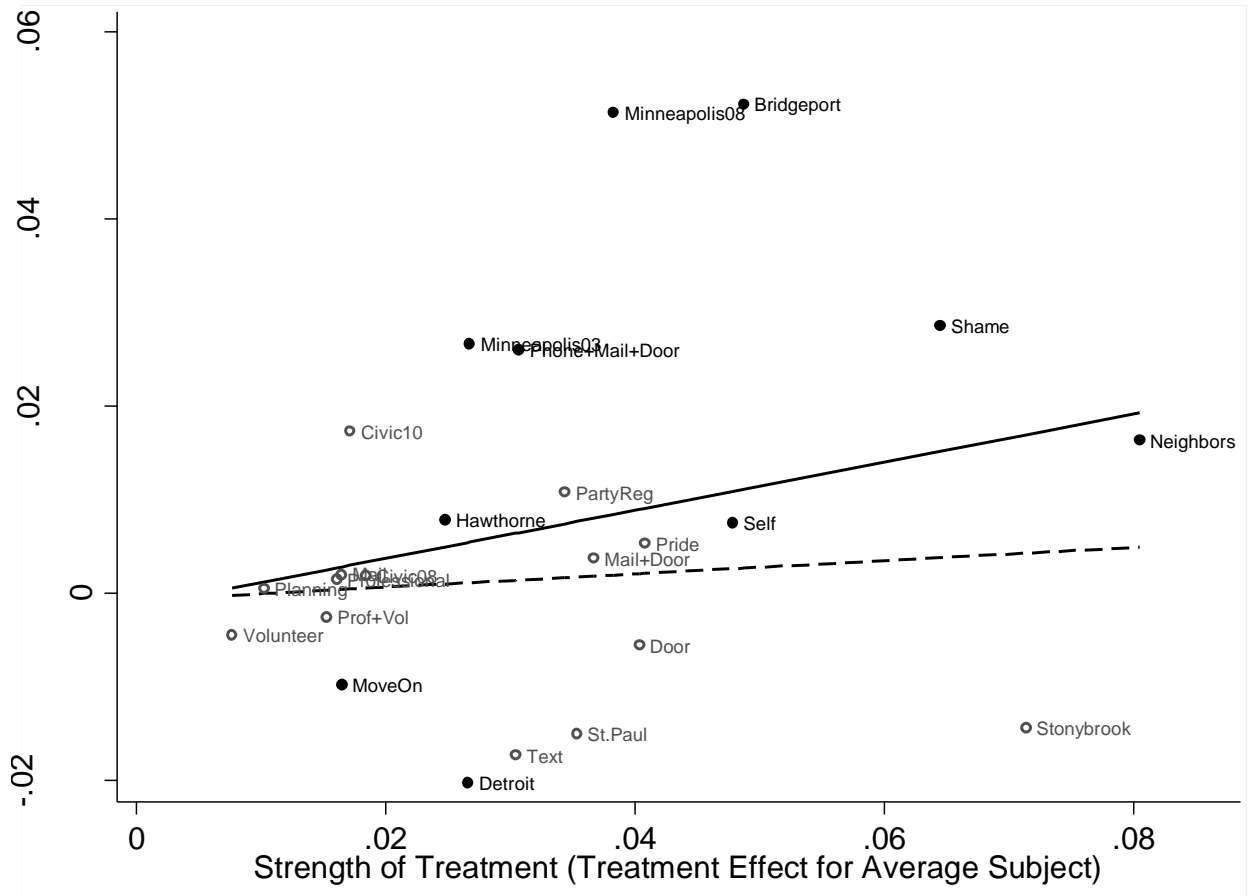
(2010). In these studies, Black citizens do appear to respond differently to GOTV efforts than non-Black citizens. In two studies, we find that high-propensity Blacks are actually *demobilized* by the treatment and in the other they were mobilized much less than whites or low-propensity Blacks. If GOTV efforts are demobilizing likely African American voters, this undermines the purpose of many efforts, in addition to presenting serious ethical concerns. Of course, this is a preliminary analysis on only three existing data sources so we draw no strong conclusions about these mechanisms. Moreover, we would expect to see some negative interactions by chance alone, so we should not draw strong conclusions from the few cases where we see this.

Figure A1. Variation across Electoral Salience



The figure assesses the salience hypothesis of Arceneaux and Nickerson (2009) that GOTV treatments will mobilize high-propensity citizens in low-salience elections and low-propensity citizens in high-salience elections. The y-axis is the multiplicative coefficient from these analyses, indicating the extent to which the treatment effect changes as propensity increases. The x-axis is the average level of turnout in the control group, a proxy for the salience of the election. Black, solid circles denote cases where the interactive coefficient is statistically significant ($p < .05$) and gray, hollow circles denote cases where the interactive coefficient is not statistically significant. The solid line represents a linear fit where all studies are weighted equally, and the dashed line indicates a linear fit where the studies are weighted by their sample sizes. The Arceneaux/Nickerson hypothesis is largely confirmed: as salience increases the effectiveness of GOTV treatments for low-propensity citizens increases relative to high-propensity citizens. However, significant variation remains in the data, suggesting that some treatments may be more or less effective in reducing the participation gap. Moreover, this analysis predicts that GOTV treatments will tend to exacerbate the participation gap in any setting where the average level of turnout is less than 50% -- essentially all electoral settings in the U.S. outside of presidential races.

Figure A2. Variation across Strength of Treatment



The figure presents the regression results from Table 3 graphically. The y-axis is the multiplicative coefficient, indicating the extent to which the treatment effect changes as the propensity variable increases. The x-axis is the additive coefficient, the treatment effect for the average citizen in the sample. Solid circles indicate that the interactive coefficient is statistically significant ($p < .05$). As before, the solid line represents a linear fit with all experiments weighted equally and the dashed line indicates a linear fit where experiments are weighted by sample size. We see that the interactive effect tends to be larger for experiments with larger average effects.

VII. More Detailed Results

The following graphs present the non-parametric results of the paper in more detail. For each experiment (or for each set of experiments that share a single control group), we present three plots. The first is a histogram representing the distribution of the propensity variable. The second are kernel regression of turnout and the propensity variable for the control and treatment groups, similar to the top panel of Figure 2. The final plot represents the difference between the kernel regressions for the treatment group(s) and the control group, indicating the conditional average treatment effect at each level of propensity, similar to the bottom panel of Figure 2.

Figure A3. Gerber and Green (2003)

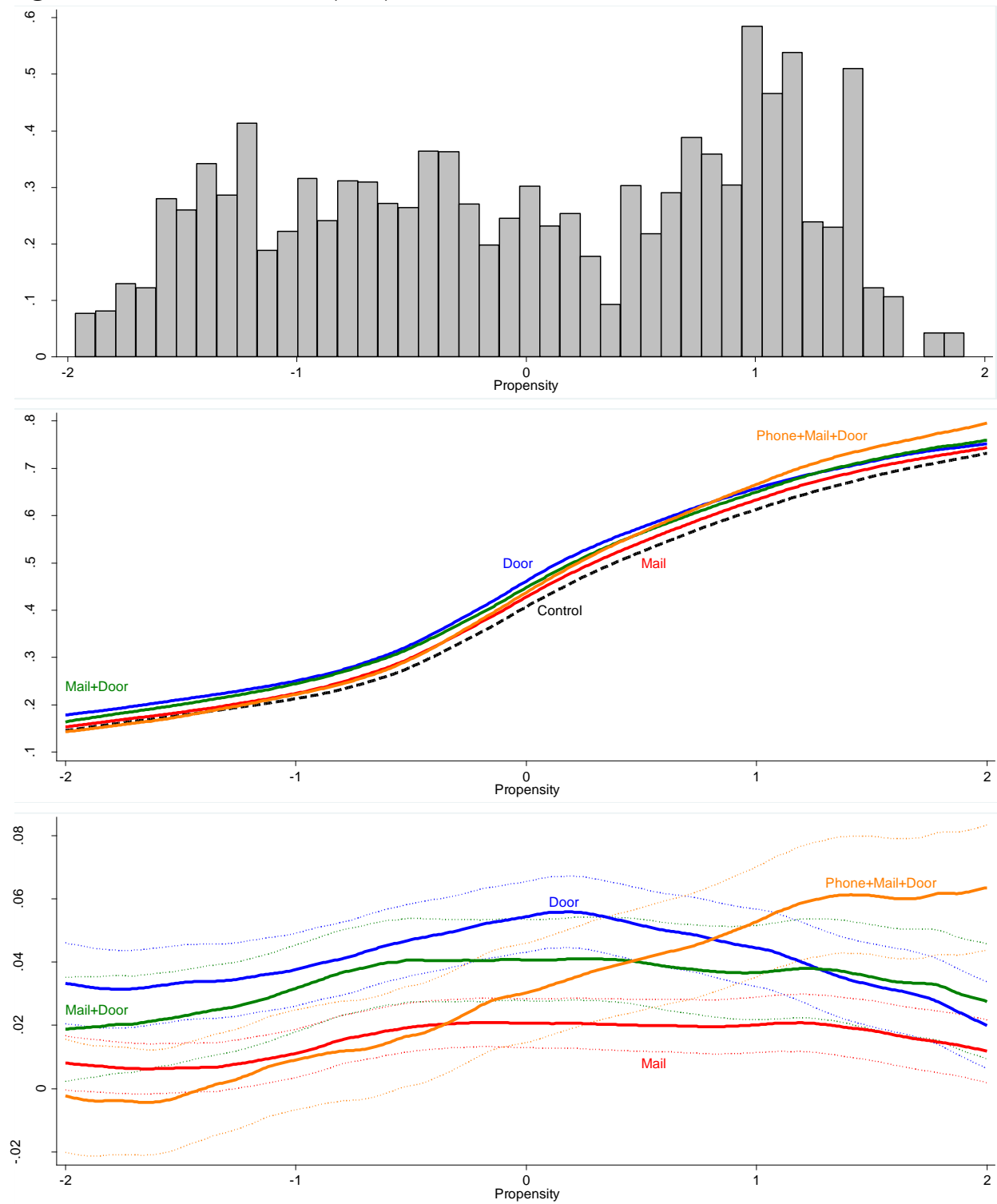


Figure A4. Bridgeport (Gerber, Green, and Nickerson 2003)

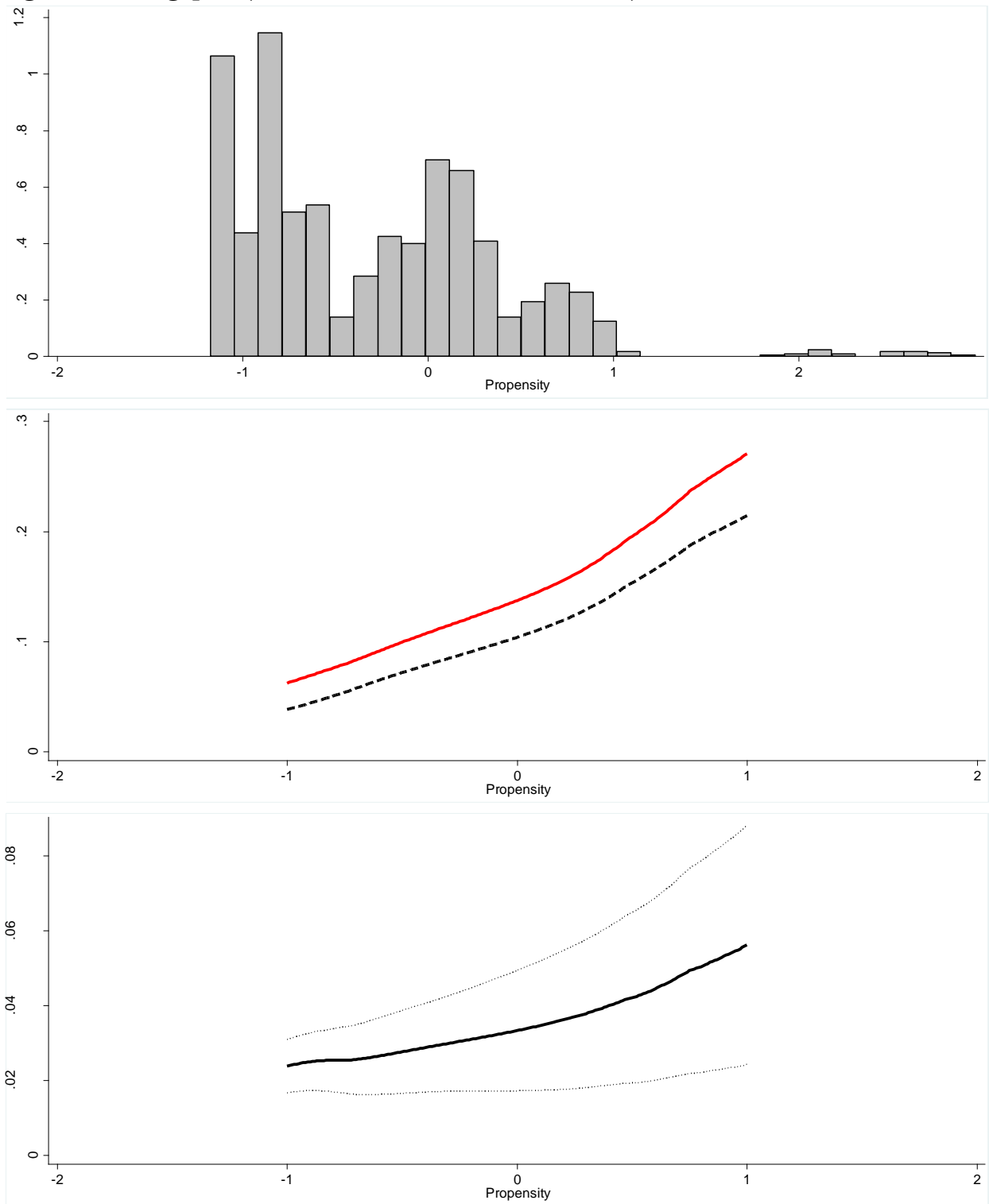


Figure A5. Detroit (Gerber, Green, and Nickerson 2003)

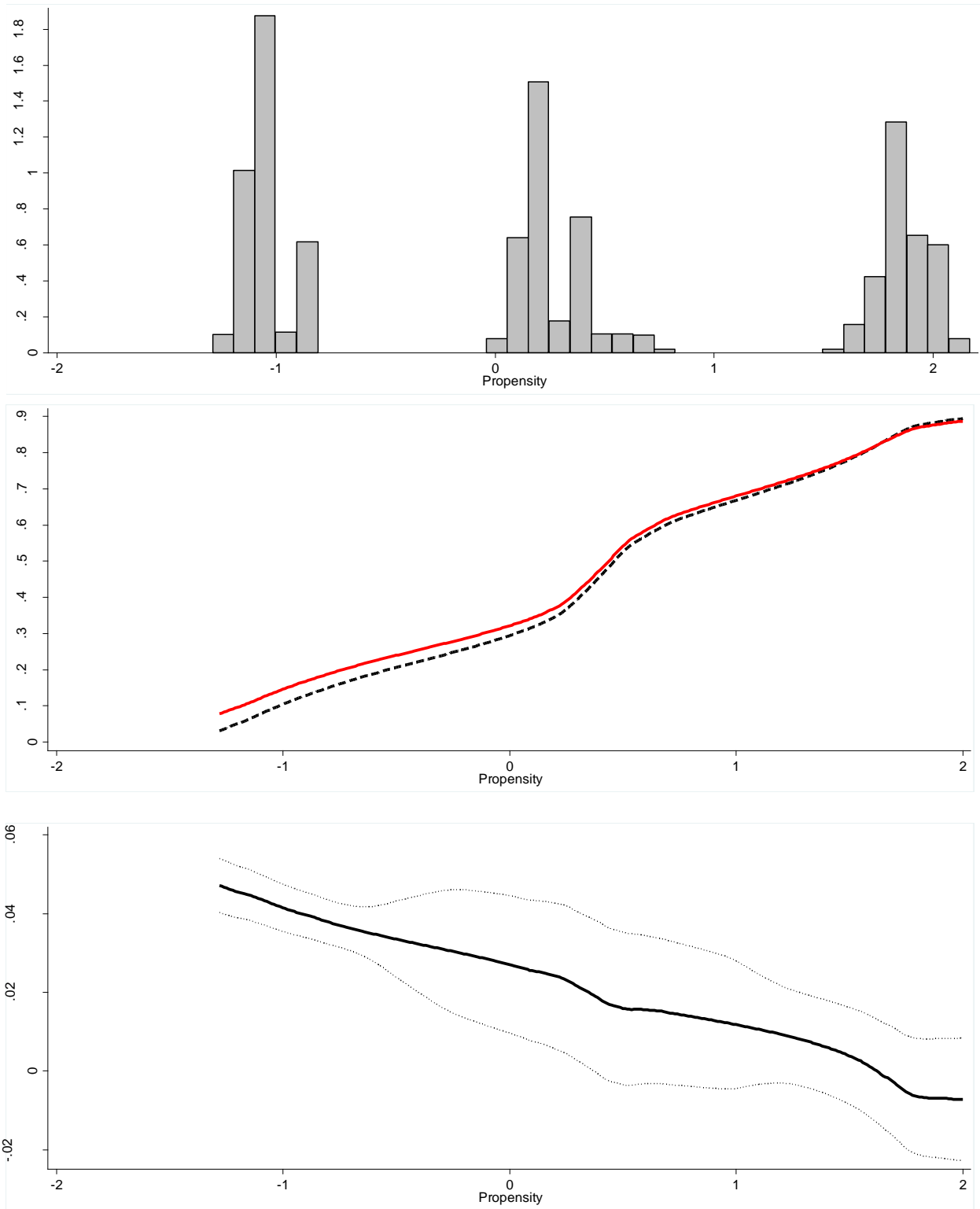


Figure A6. Minneapolis (Gerber, Green, and Nickerson 2003)

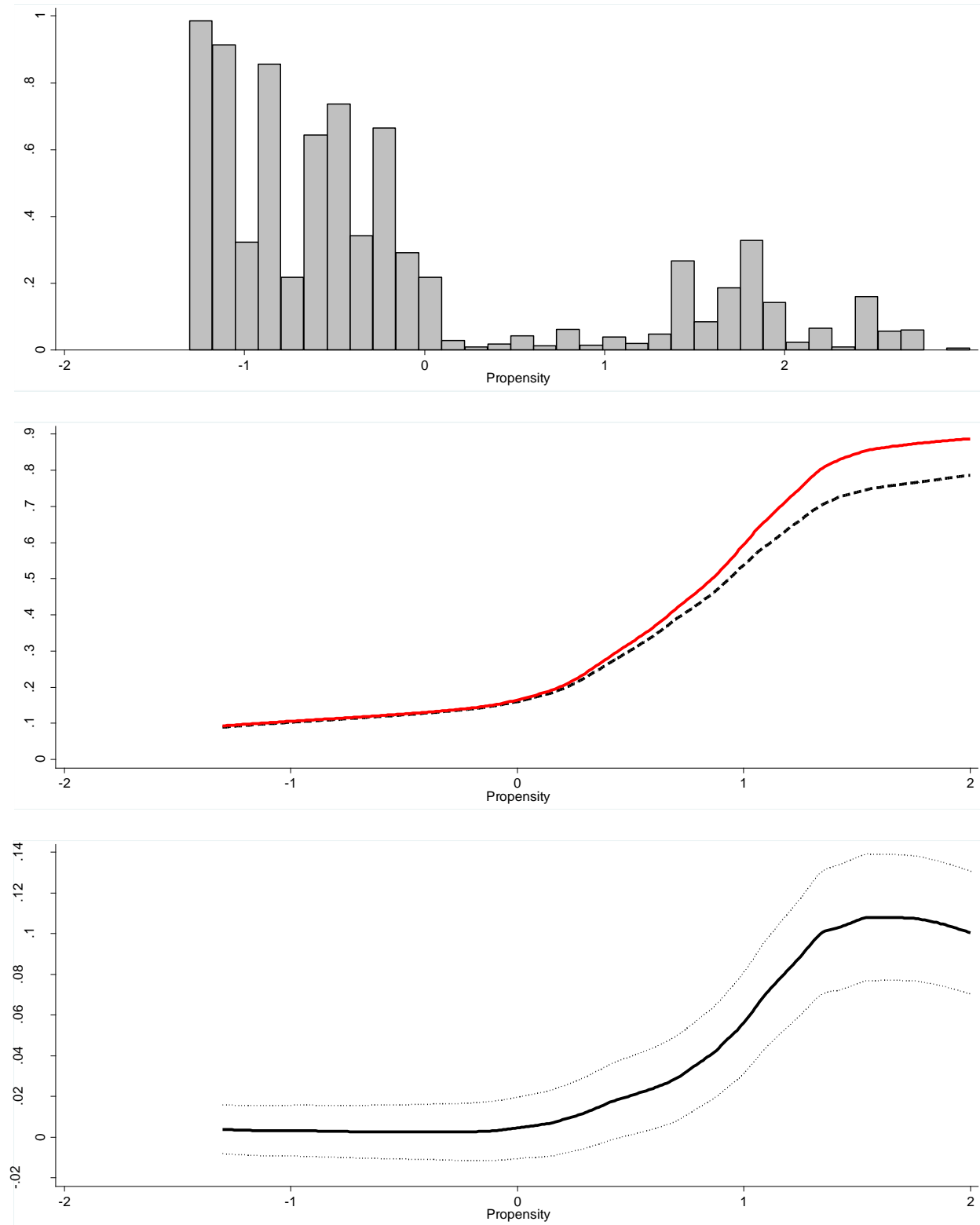


Figure A7. St. Paul (Gerber, Green, and Nickerson 2003)

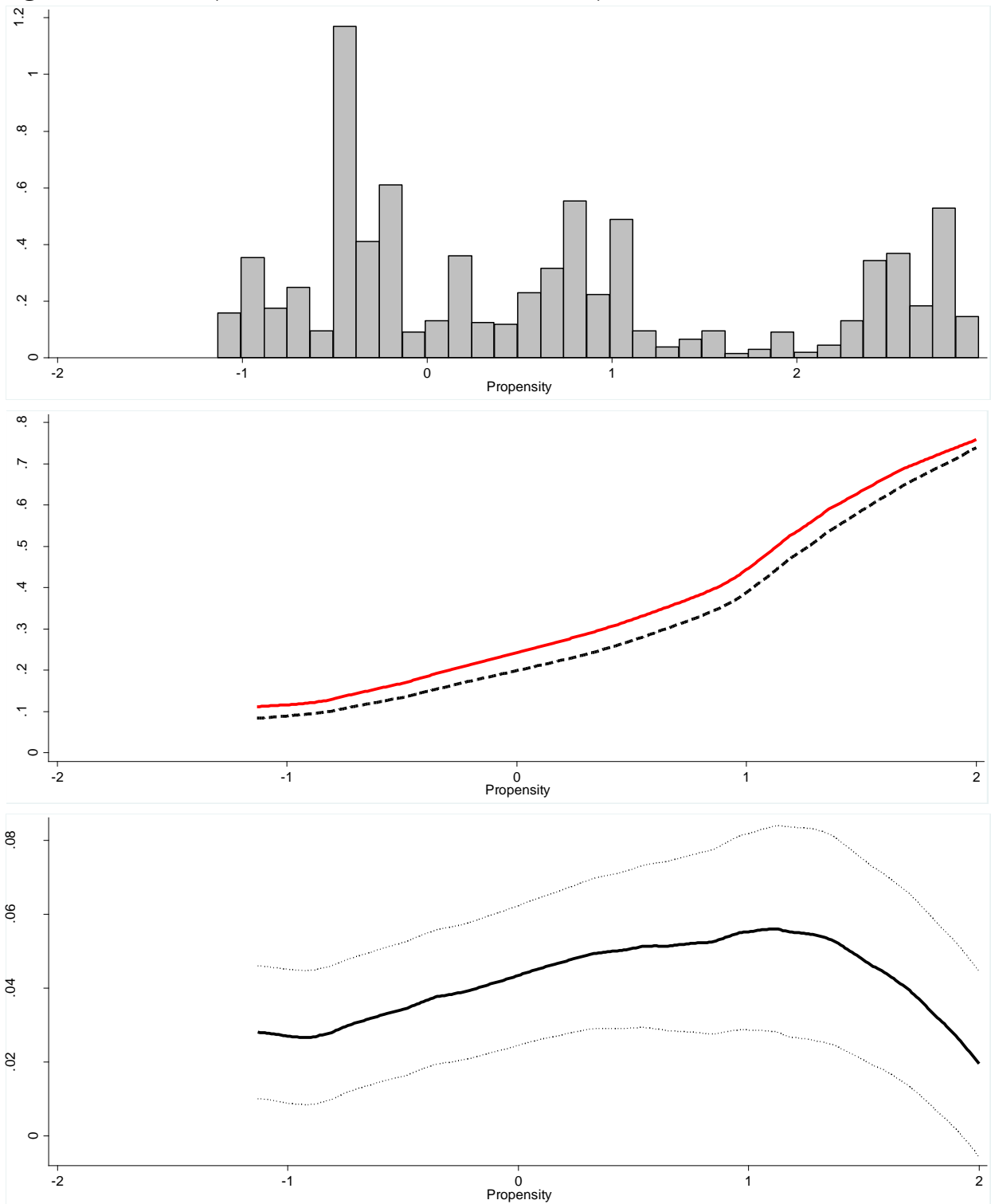


Figure A8. Stonybrook (Nickerson 2006)

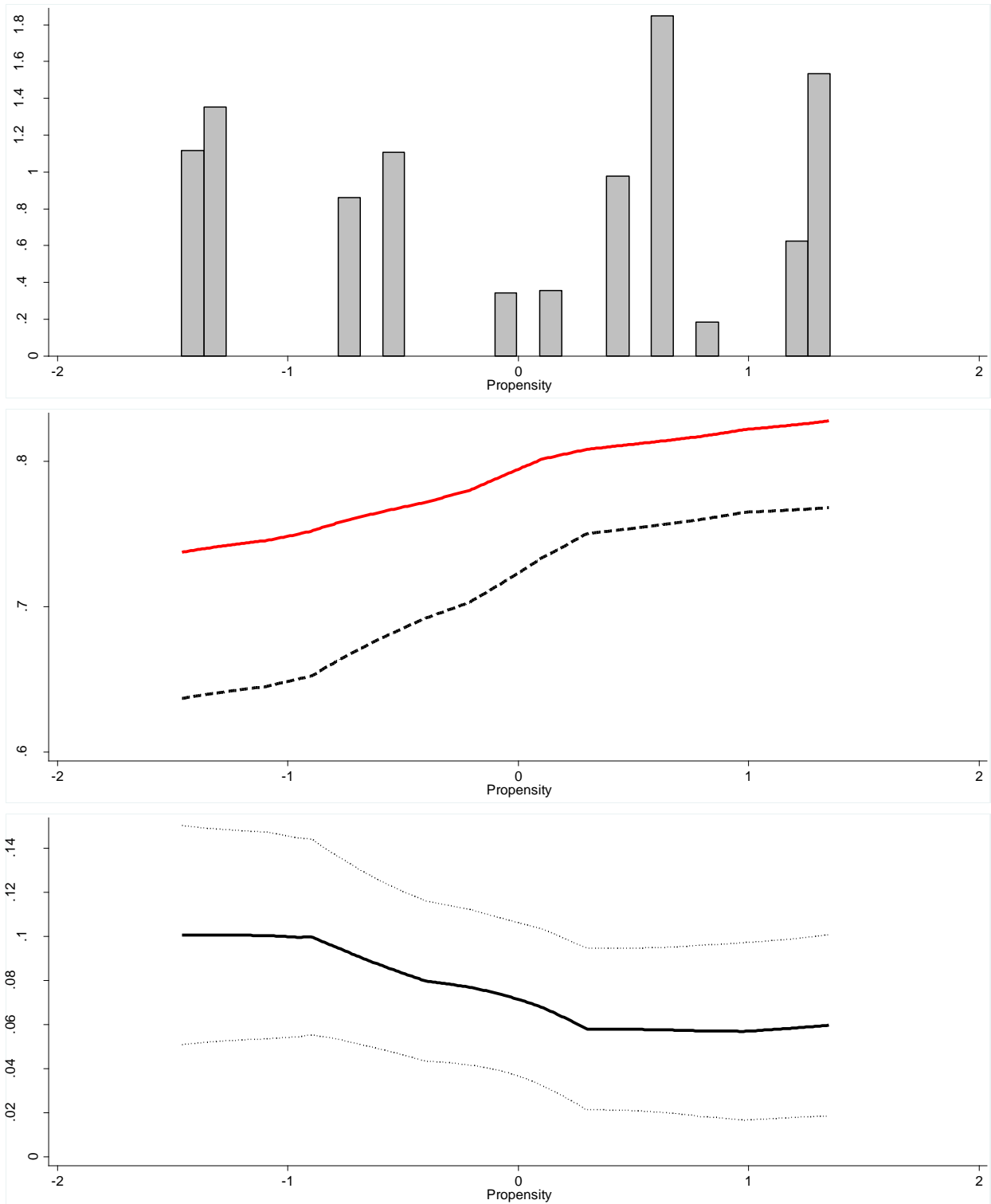


Figure A9. Nickerson (2007)

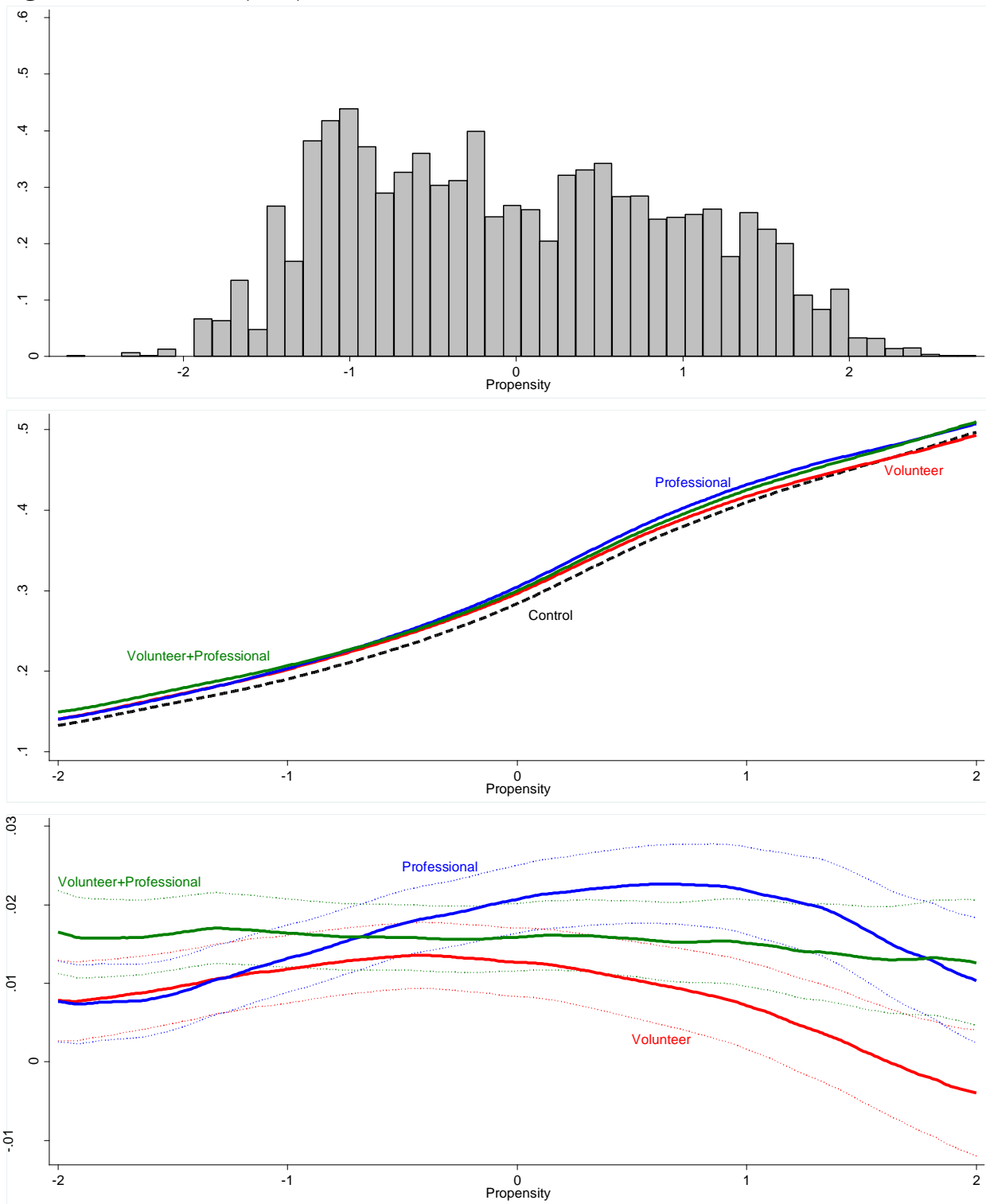


Figure A10. Gerber, Green, and Larimer (2008)

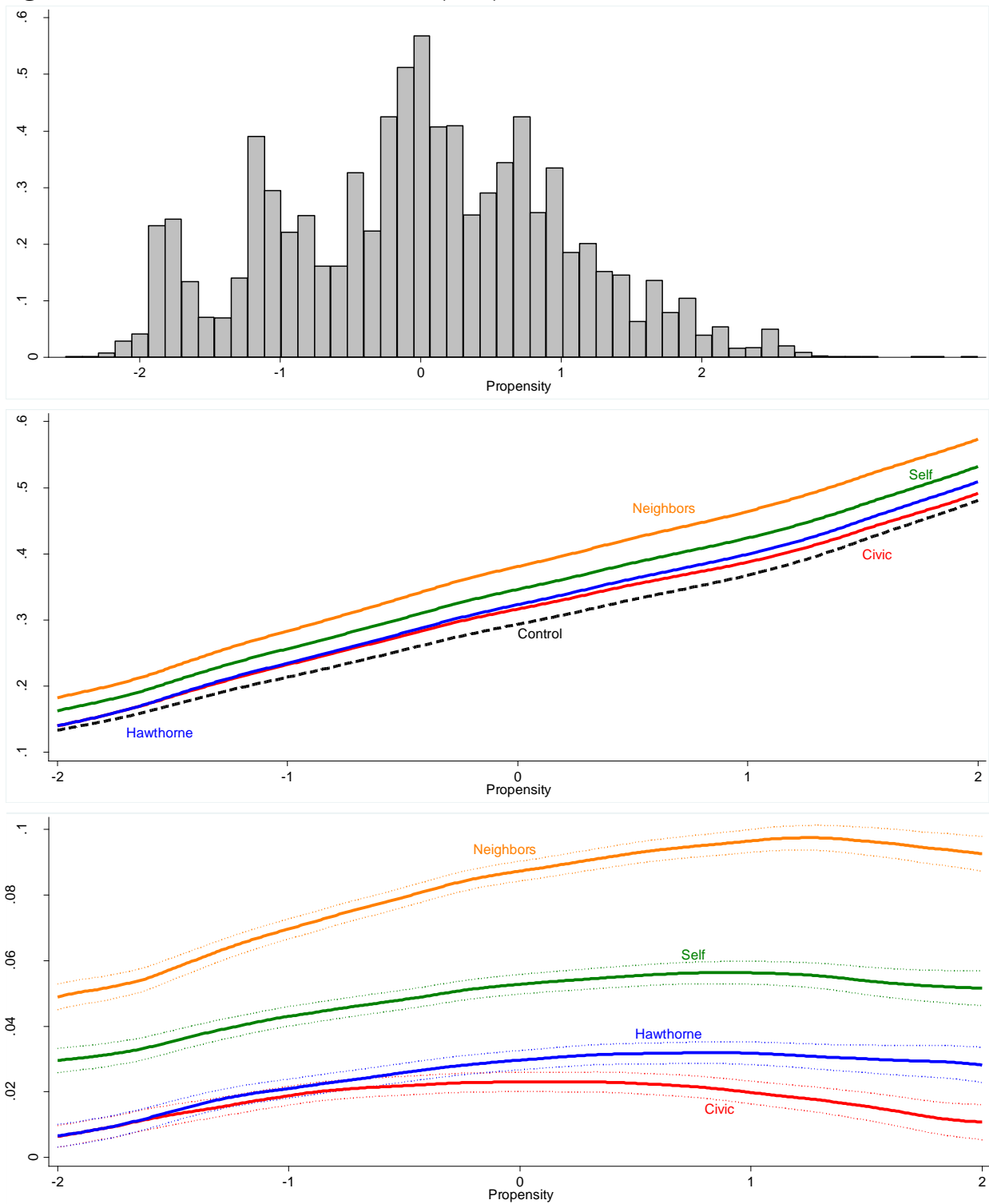


Figure A11. Move On (Middleton and Green 2008)

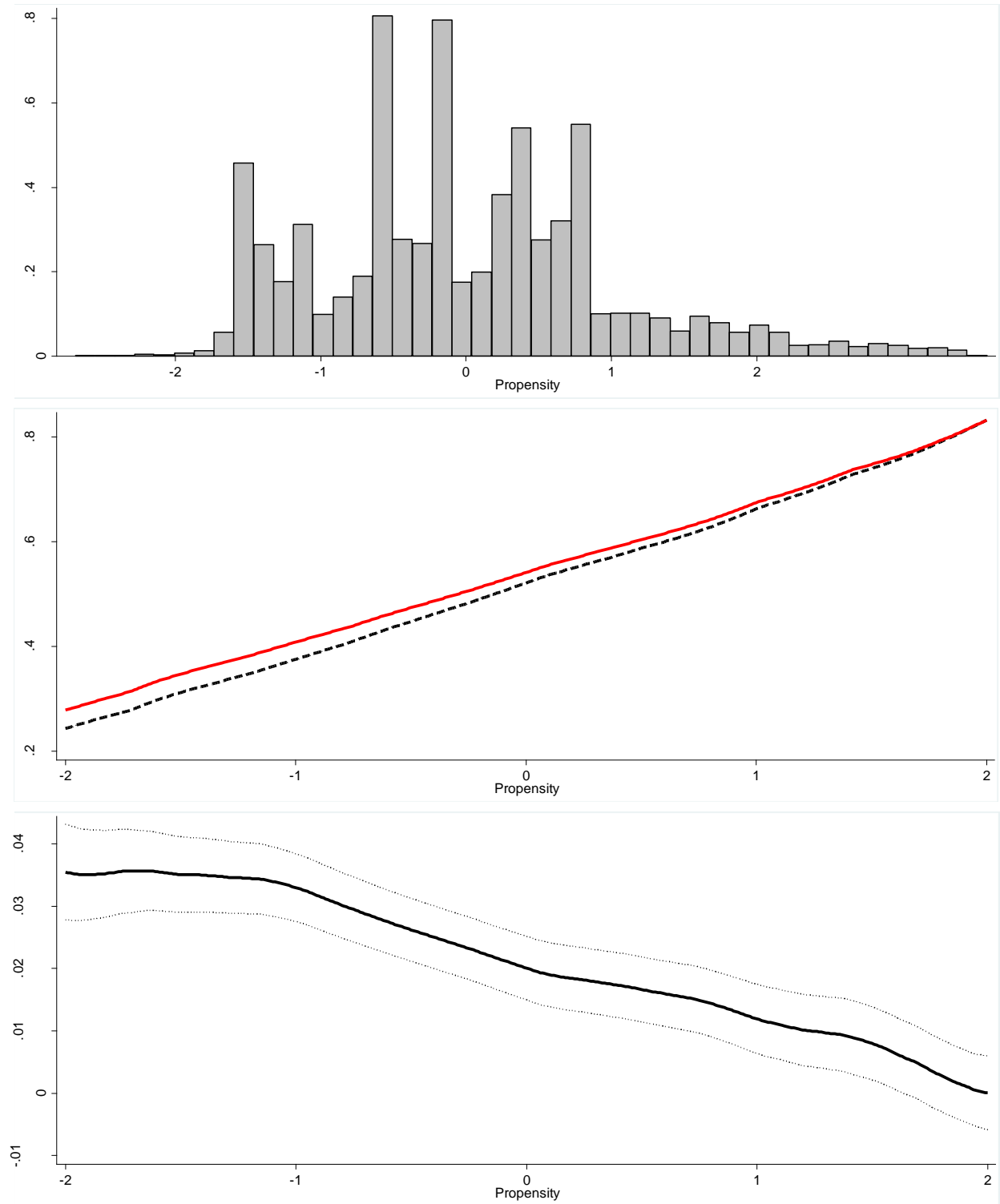


Figure A12. Minneapolis (Nickerson 2008)

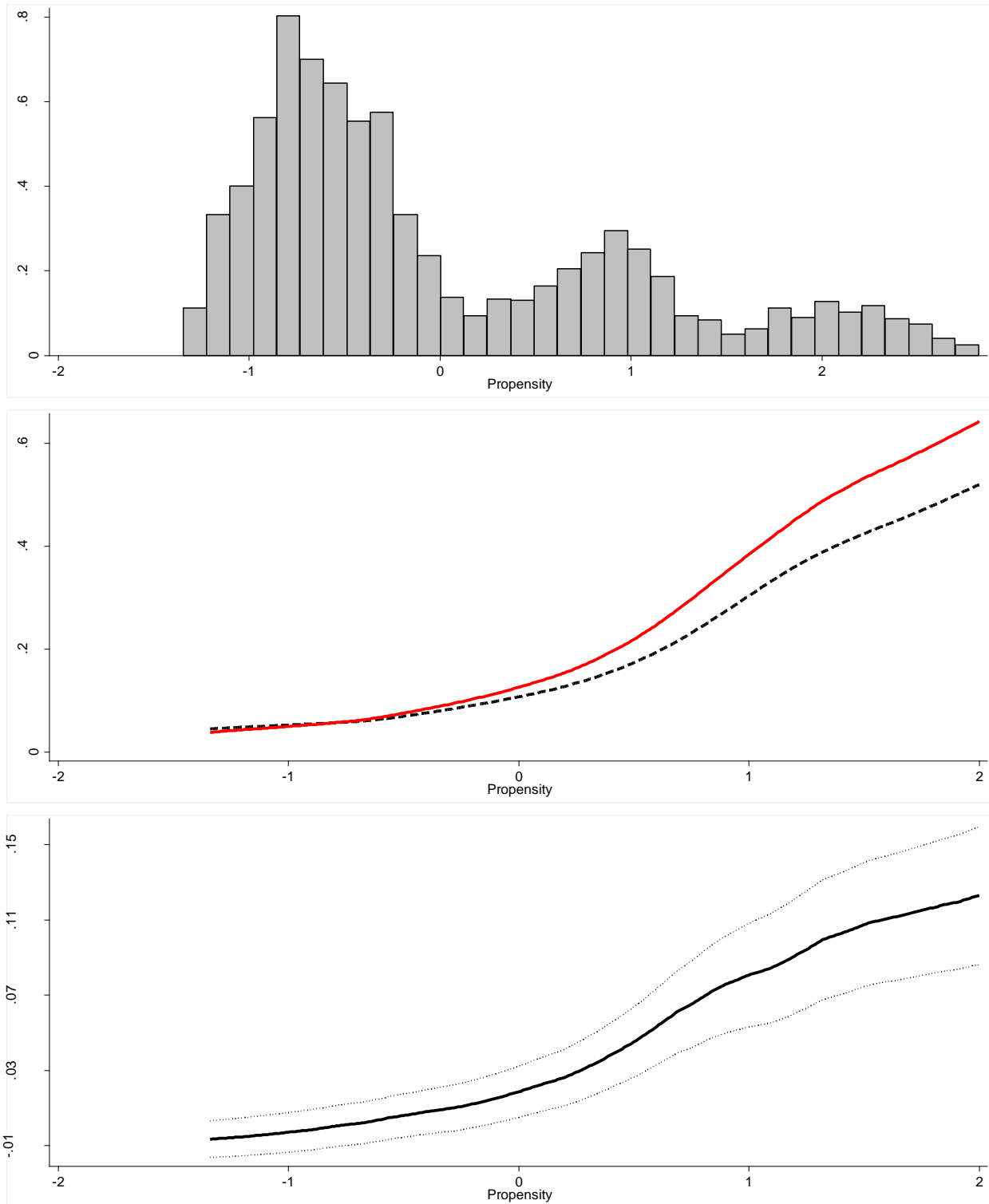


Figure A13. Text Message (Dale and Strauss 2009)

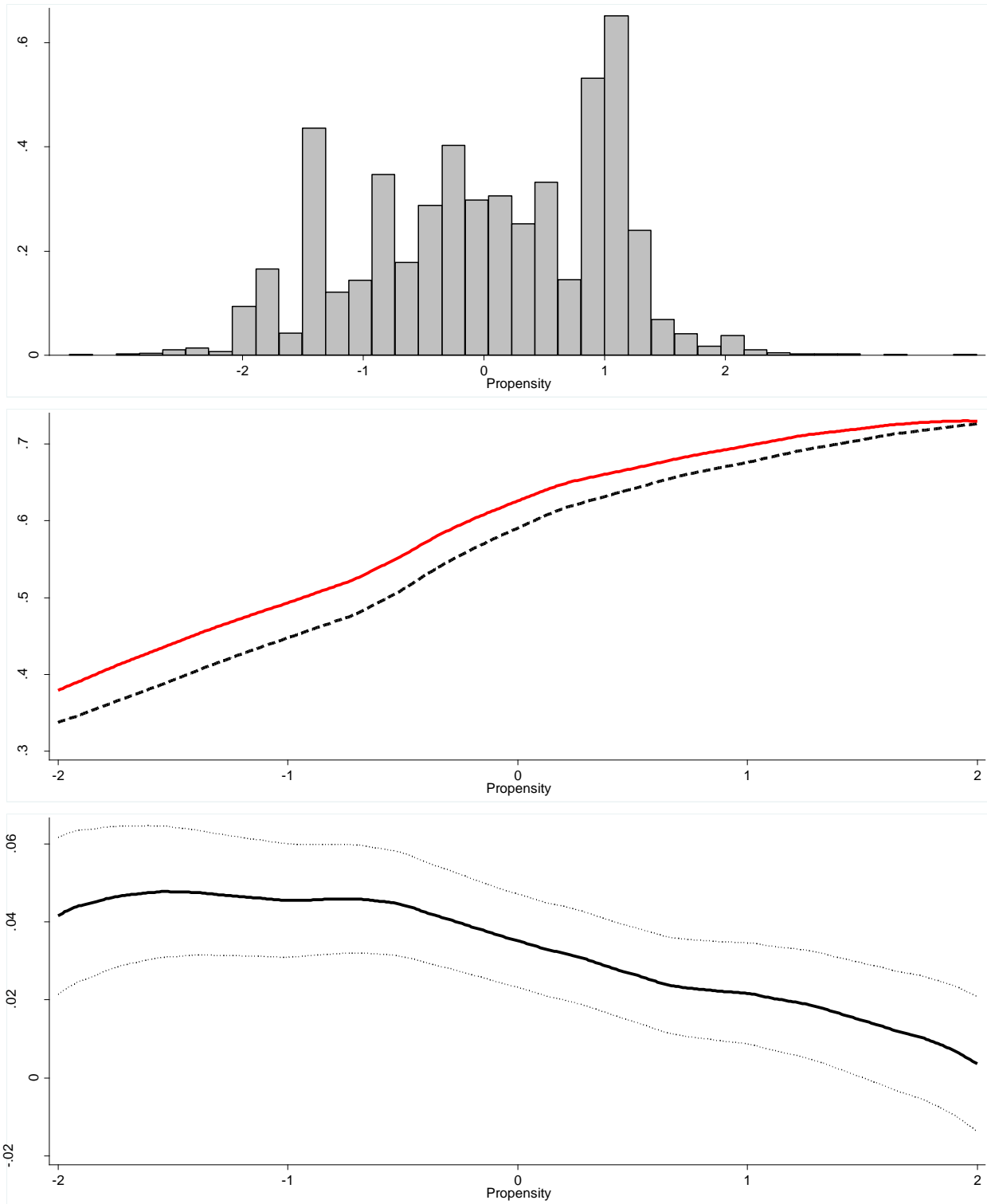


Figure A14. Gerber, Green, and Larimer (2010)

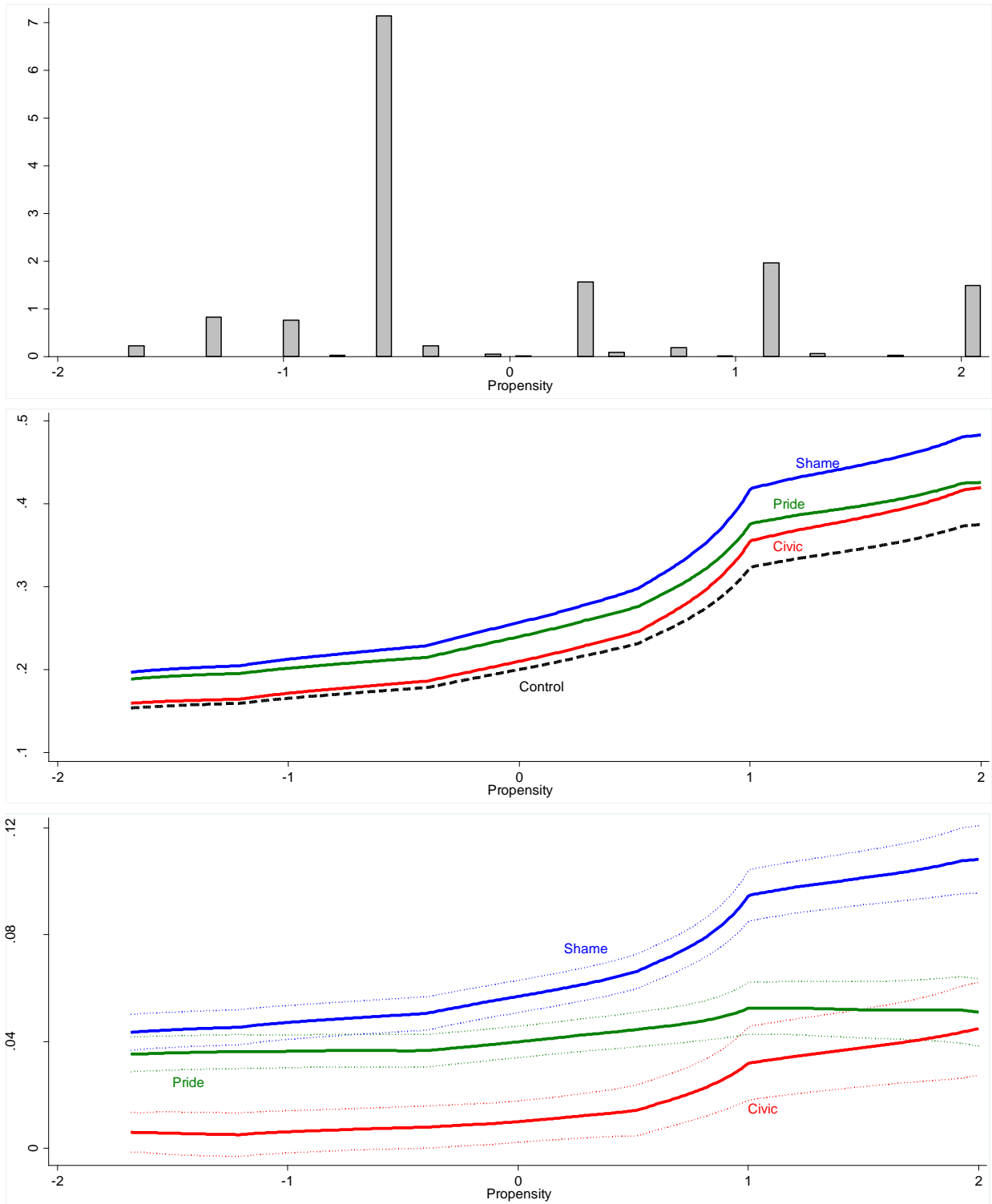


Figure A15. Party Registration (Gerber, Huber, and Washington 2010)

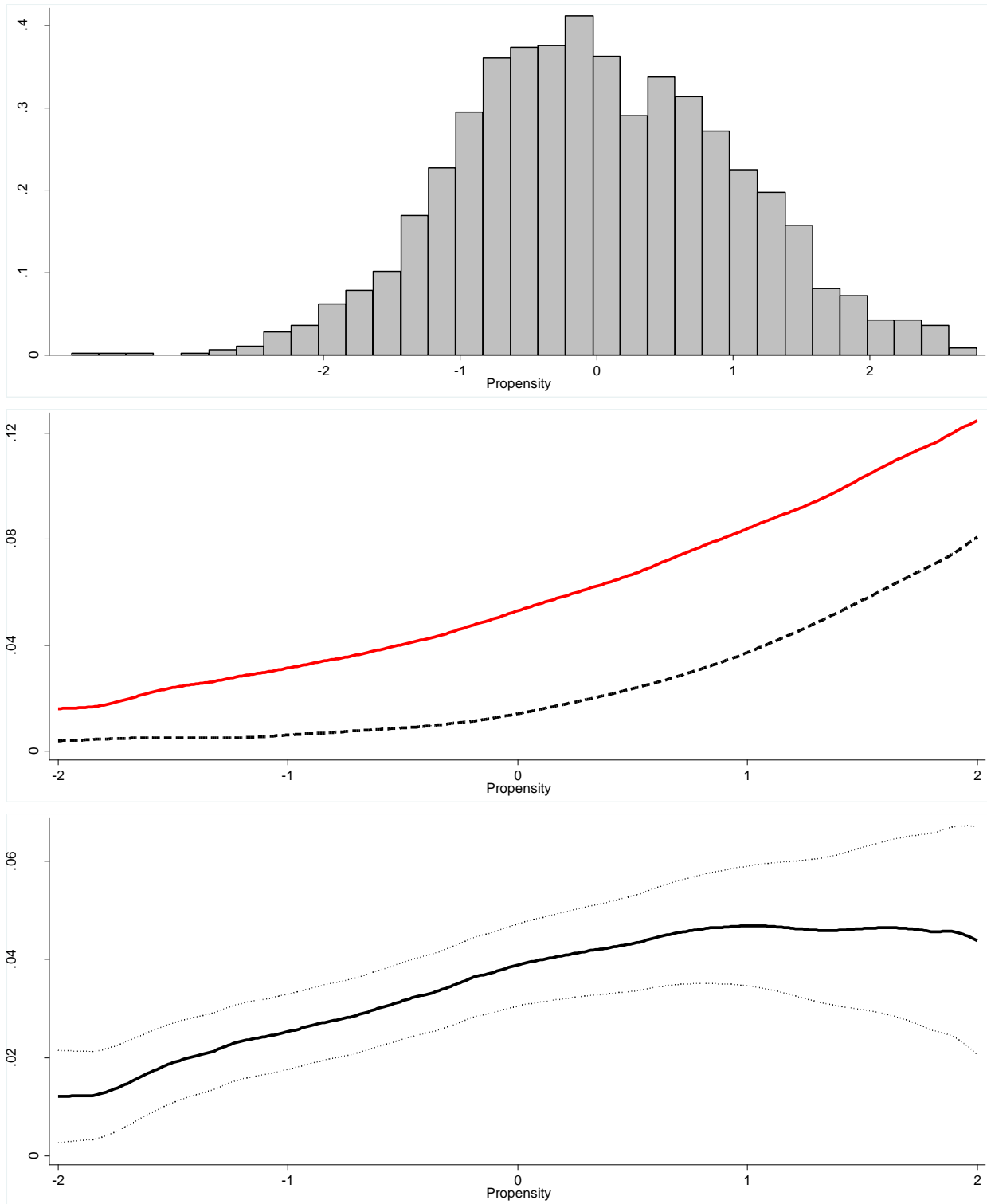
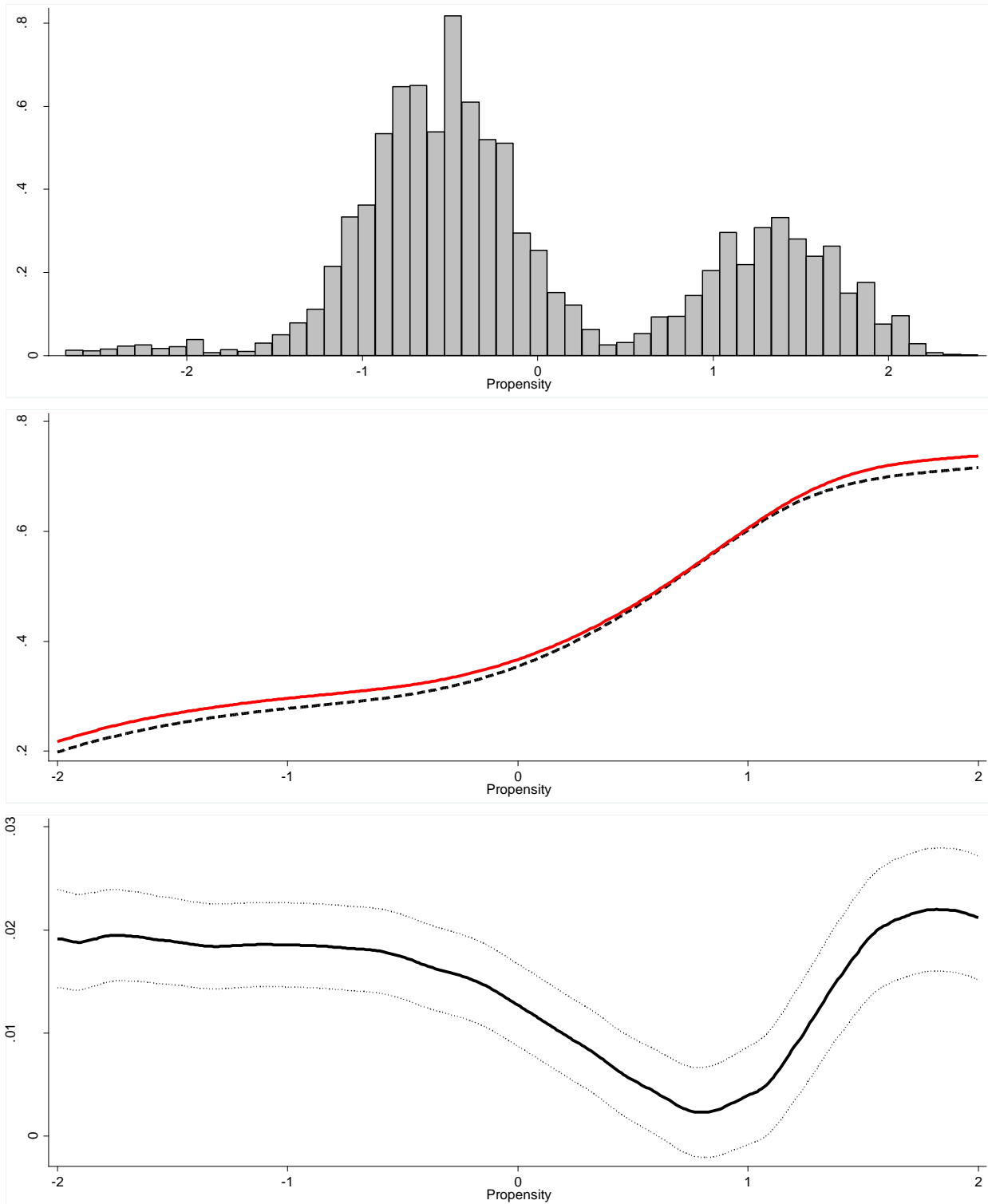


Figure A16. Planning (Nickerson and Rogers 2010)



Reference

Ansolabehere, Stephen, Eitan Hersh, Alan Gerber, and David Doherty. 2010. Voter Registration List Quality Pilot Studies. *Caltech/MIT Voting Technology Project*
<http://vote.caltech.edu/drupal/node/335>.