Voting to Tell Others*

Stefano DellaVigna
UC Berkeley and NBER

John A. List
U Chicago and NBER

Ulrike Malmendier
UC Berkeley and NBER

Gautam Rao
Microsoft Research and Harvard University

This version: April 16, 2015

Abstract

Why do people vote? We design a field experiment to estimate a model of voting 'because others will ask'. The expectation of being asked motivates turnout if individuals derive pride from telling others that they voted, or feel shame from admitting that they did not vote, provided that lying is costly. In a door-to-door survey about election turnout, we experimentally vary (i) the informational content and use of a flyer pre-announcing the survey, (ii) the duration and payment for the survey, and (iii) the incentives to lie about past voting. The experimental results indicate significant social image concerns. For the 2010 Congressional election, we estimate a value of voting 'because others will ask' of \$5-\$15, one of the first estimates of the value of voting in the literature.

*We thank Nageeb Ali, Leonardo Bursztyn, David Card, Raj Chetty, Justin Fox, Andrew Gelman, Alan Gerber, Yosh Halberstam, Mitch Hoffman, Greg Huber, Tom Palfrey, Ricardo Perez-Truglia, Todd Rogers, Glen Weyl, Noam Yuchtman, audiences at Berkeley, Boston University, Harvard, Microsoft Research, RAND, Stanford, Toulouse, UCSD, and at the 2011 ESA World Congress (Chicago), the 2012 North-American ESA Tuscon Conference, the 2012 NBER Political Economy Summer Institute, the 2013 Yale CSAP summer workshop, and the 2013 SITE Psychology and Economics conference for helpful comments. We also thank Alec Brandon, Tristan Gagnon-Bartsch, Sheng Li, David Novgorodsky, Jessica Shui, and Vera te Velde for excellent research assistance.

1 Introduction

Get-out-the-vote interventions in the spirit of Gerber and Green (2000) have attracted significant attention by researchers and practitioners alike. In the most effective intervention to date, sending a letter about the turnout record of neighbors increases turnout by 8 percentage points (Gerber, Green and Larimer, 2008). This remarkable effect likely reflects social-image concerns: the explicit threat to make their voting record public.

We consider a related, but more commonplace, social image motivation for voting. While it is rare for others to confront us with our voting record, it is common for neighbors, friends, and family to ask whether we voted. If individuals care about what others think of them, they may derive pride from telling others that they voted or feel shame from admitting that they did not vote. In addition, they may incur disutility from lying about their voting behavior.

Such individuals are motivated to vote (also) because they anticipate that others will ask if they did. If they vote, they can advertise their 'good behavior' when asked. If they do not vote, they face the choice of being truthful but incurring shame, or saying that they voted but incurring the lying cost. This trade-off is reflected in the fact that 25 to 50 percent of non-voters lie when asked about their past turnout (Silver, Anderson, and Abramson, 1986).

In this paper, we estimate this model of voting 'to tell others', which follows Harbaugh (1996), using a natural field experiment. Due to the tight link between the model and the experiment, we are able to estimate the value of voting that is due to this social image motivation. This is a significant contribution given the rarity of estimates of the value of voting in the political economy literature.

The main experiment took place in the summer and fall of 2011 in the suburbs of Chicago. We visited households and asked whether they were willing to answer a short survey, including a question on whether they voted in the 2010 congressional election. In some cases, we posted a flyer on the doorknob a day in advance to announce the upcoming survey. Unbeknownst to the households, we used voting records to restrict the sample to households where either all registered members voted in the 2010 elections (henceforth, voting households) or none of the registered members voted in 2010 (non-voting households). We did not visit households with a mixed 2010 voting record.

The field experiment has three main sets of treatments. In the first set, we randomize the information on the flyer. In one group, the flyer informs households that the next day we will ask for their participation in a door-to-door survey. In a second group, the flyer specifies that the survey will be about "*your voter participation in the 2010 congressional election.*" Changes in the share of households opening the door and completing the survey between the first and the second group reflect the value of being asked about voting. An increase in the participation of voting households indicates the pride of saying that they voted. A decrease among non-voting households indicates shame from admitting that they did not vote and a cost of lying.[1]

We find that, on average, voters do not sort in. In fact, voting households are slightly less likely to answer the door and do the survey when they are informed about the turnout question. Non-voters sort out significantly, decreasing their survey participation by 20 percent.[2]

These results may depend on the particular election considered. The 2010 elections were disappointing for Democrats and positive for Republicans, including in Illinois the loss of President Obama's previous seat in the Senate. The lack of pride among voters may reflect disappointment, given that the neighborhoods visited were largely Democratic. Indeed, if we restrict the analysis to voters registered for the Republican primaries, we find evidence of sorting in.

The findings on sorting provide prima facie evidence of social-image utility. In order to quantify the utility value, we measure the cost of sorting in and out of answering the survey. To do so, we introduce a second set of (crossed) randomizations, in which we vary the promised payment for the survey ($10 versus $0) and the pre-announced duration (5 minutes versus 10 minutes). We find that the effect of reducing payment by $10 is comparable to the sorting response of non-voters to the election flyer, implying significant social-image (dis)utility.

To estimate the value of voting '*because others will ask*,' we need additional counterfactual social-image values, such as the shame that voters would feel were they to say they did not vote. These counterfactuals are not provided by the sorting moments.

We thus introduce a third set of crossed treatments. We randomize incentives to provide a different response to the turnout question. Specifically, we inform half the respondents of the ten-minute survey that the survey will be eight minutes shorter if they state that they did not vote in the 2010 congressional election. For voters, this treatment amounts to an incentive to lie and permits us to quantify the disutility of voters were they to say (untruthfully) that they did not vote. For the 50 percent of non-voters who lie, this treatment provides an incentive to tell the truth. We provide a parallel $5 incentive in the 5-minute survey to state that one did not vote.

This novel experimental design makes it possible to price out how much respondents care about making a particular statement. This approach has applications also to other settings where responses could have social image or signalling motivations, such as in contingent valuation surveys, or surveys of sensitive political and social attitudes.

The results reveal that non-voters are significantly more sensitive to these incentives than voters. When incentivized, the share of non-voters who lie decreases significantly, by 12 per-

[1]This randomization also includes a group with no flyer, as well a group with an opt-out box.

[2]We also randomize the information provided by the surveyor at the door. For half of the households, they indicated a survey "on your voter participation in the 2010 congressional election." This manipulation, which was crossed with all other manipulations, did not have a significant effect on survery take-up for either voters or non-voters.

centage points, while the share of voters who lie increases only insignificantly, by 2 percentage points. The results are similar for time and monetary incentives, and reveal a strong preference of voters for saying that they voted.

We combine the moments from the three sets of treatments to estimate the parameters of our model using a minimum-distance estimator. The benchmark estimates provide no evidence of pride in voting. On average, voters get negative utility from saying that they voted. However, voters obtain an even lower utility, by \$15, from untruthfully saying that they did not vote. Non-voters are estimated to be on average indifferent between saying truthfully that they did not vote or lying and saying that they voted, with negative average utility from either option. We estimate substantial heterogeneity in social image utility, especially among voters.

These estimates identify the key parameters up to an additive lying cost, which remains unidentified. Since the lying cost is an integral part of the social-image value of voting, we adopt two approaches to address this limitation. First, we identify a subsample of households with similar turnout histories prior to 2010, but different turnout in 2010. Voters and non-voters (in 2010) in this subsample are likely to be similar, and we assume that they have the same social-image and lying parameters. Under this assumption, we estimate lying costs of \$5, leading to a value of being asked once of \$1.50. Second, we compute the value of voting for a range of values of the lying cost, including one we estimate from the laboratory evidence in Erat and Gneezy (2012). In this range, the implied value of voting 'because of being asked *once*' is in the range of \$1-\$3 for voters. Hence, the estimates are quite similar under both approaches.

To compute the overall value of voting due to being asked, we scale up the estimated value of being asked *once* by the average number of times asked. Our survey respondents report being asked, on average, five times whether they voted in the 2010 congressional election, implying an estimated value of voting 'because others will ask' of \$8 for the subsample of voters, and in the range of \$5-\$15 for the considered range of lying costs. This sizeable magnitude likely understates the value of voting 'to tell others', since it is based on being asked by a (previously unknown) surveyor. The social-image utility and the lying cost from interactions with family, friends, and co-workers are likely to be larger.

Two implications are worth emphasizing. First, for a reasonable range of lying costs, the value of voting 'because others ask' is larger for voters than for non-voters, consistent with cross-sectional differences in turnout. Second, while the estimates are based on a congressional election, our survey respondents report being asked nearly twice as often about voting in presidential elections. Under the (strong) assumption of similar social-image values, the value of voting 'to tell others' in presidential elections is about twice as high, in the range of \$10-\$30, consistent with the observed higher turnout in presidential elections.

The main field experiment was designed to measure the value of voting without affecting voting itself – a crucial difference from the get-out-the-vote literature. Instead, we rely on sorting, survey completion, and survey responses. This allows us to estimate the magnitudes and signs of the social image utility associated with being asked about voting (a common occurrence). But the model also suggests an obvious intervention to increase turnout: individuals with social-image motives are more likely to vote the more they expect to be asked. Experimentally increasing this expectation should thus lead to an increase in turnout.

In November of 2010 and of 2012, we did just that. A few days before the election, a flyer on the doorknob of treatment households informed them that '*researchers will contact you within three weeks of the election [...] to conduct a survey on your voter participation.*'[3] A control group received a flyer with a mere reminder of the upcoming election. The results are consistent with the model, though statistically imprecise. In 2010, the turnout of the treatment group is 1.3 percentage points higher than the control group (with a one-sided p-value of 0.06). In the higher-turnout presidential election of 2012, the turnout difference is just 0.1 percentage points (not significant). The results are consistent with the contemporaneous results of Rogers and Ternovski (2013), who also inform a treatment group that they may be called after the election about their voting behavior, and also find a positive impact on turnout. The much smaller effect sizes than in Gerber, Green and Larimer (2008) are not surprising, since they estimate the effect of informing neighbors about the official turnout record, while we isolate the effect of (at most) one more interaction with a questioner, where lying remains an option.[4]

Are the get-out-the-vote results consistent with the estimates of the value of voting? Using the model estimates, we predict the impact on the turnout decision of an announced visit under two alternative assumptions about the variance of the other reasons to vote (which we do not observe). The observed turnout effect in 2010 is consistent with the model for lying costs in the range of \$2-\$10 (depending on the assumed variance). Further, we predict a smaller effect for presidential elections, holding the social image parameters constants, also consistent with the data. Thus, the get-out-the-vote results provide a validation of the estimates, albeit a somewhat imprecise one.

Finally, we would like to mention some caveats and alternative interpretations. First, the results are specific to their time and location—the 2010 congressional elections in Illinois. As we discussed, the lack of estimated pride in voting is possibly related to the disappointing results for Democrats in 2010, and could more generally be a function of the aggregate turnout rates, closeness and importance of the election. It will be interesting to apply this methodology to other elections to test this directly in future work.

Second, we address the important concern that the observed 'sorting out' among non-voters may reflect a dislike of talking about politics, independent of their non-participation in the election. When we allow for a different taste among voters and non-voters for talking about

[3] We follow up with a door-to-door visit, as advertised.

[4] In addition, Gerber, Green, and Larimer (2008) focus on non-competitive primary elections, where turnout interventions lead to larger effects than for competitive general elections like the ones we study.

politics, we lose the ability to estimate one of the social-image parameters. But the net value of voting 'due to being asked' is still identified and in fact remains unchanged, since it is identified by the lying treatments. Intuitively, while a differing taste for talking about politics could explain the sorting patterns in response to the flyer treatments, it does not explain the lying about voting, nor the differential response of voters and non-voters to the lying incentives.

Third, in a series of robustness and sensitivity checks, we relax and vary numerous assumptions of the model, including allowing for measurement error in the voting record, dropping various groups of moments, assuming correlated social image parameters, varying the elasticity of sorting, etc. The estimated value of voting remains largely robust to these variations.

Fourth, the estimation of the parameters relies on the full set of 100 moments from the field experiment, making it difficult to highlight which moments play the most important role for the key parameter estimates. To address this issue, we plot the Gentzkow and Shapiro (2013) sensitivity measures, highlighting the critical role of the lying interventions and confirming the intuition we provide.

In addition to complementing the substantial literature on get-out-the-vote field experiments, summarized in Green and Gerber (2008), this paper more broadly contributes to the vast literature on why people vote.[5] Our main contribution is to provide an estimate of the value of voting (due to being asked), which is rare in the literature. Among the few papers which do so, Coate and Conlin (2004) and Coate, Conlin and Moro (2008) estimate, respectively, a group-rule utilitarian model and a pivotal-voting model on alcohol-regulation referenda data. Their estimates are up to a scaling for the voting cost, which is not identified. Levine and Palfrey (2007) estimate a pivotal-voting model, but use laboratory elections where parameters can be controlled. In contrast, we obtain estimates of the value of voting by virtue of the design of the field experiment.

The paper also relates to the literature on social image. The theoretical papers micro-found social-image concerns as signaling models (Benabou and Tirole, 2006; Andreoni and Bernheim, 2010; Ali and Lin, 2013) and suggest intriguing possibilities for how our estimated social image parameters might vary in different elections with differing degrees of turnout, closeness and importance. The empirical papers highlight the impact of social image on productivity (Ashraf, Bandiera, and Jack, 2014), contributions to public goods (Ariely, Bracha, and Meier, 2009; Lacetera and Macis, 2010), campaign contributions (Perez-Truglia and Cruces, 2013), and energy consumption (Allcott, 2011). Our study attempts to bring these literatures closer by providing estimates of the social-image parameters. We hope that future research strengthens the ties, providing estimates of the underlying signaling game.

This paper also complements a small but growing literature on behavioral political econ-

[5]This vast literature inludes Downs (1957), Ledyard (1984), and Palfrey and Rosenthal (1983, 1985)) on pivotal voting and Riker and Ordeshook (1968), Harsanyi (1977), Blais (2000), and Feddersen and Sandroni (2000) on norm-based voting.

omy, including Shue and Luttmer (2009), Finan and Schechter (2012), Passarelli and Tabellini (2013), and Bursztyn et al. (2014). This paper links this literature with the literature on structural behavioral economics (Laibson, Repetto, and Tobacman, 2007, Conlin, O'Donoghue, and Vogelsang, 2007; DellaVigna, List, and Malmendier, 2012).

The remainder of the paper proceeds as follows. The next section introduces the model. Section 3 summarizes the experimental design. Sections 4 and 5 present, respectively, the reduced-form results and structural estimates for the main experiment. Section 6 introduces the get-out-the-vote experiment. Section 7 concludes.

2 Model

Voting. Voting depends on four factors: pivotality, warm glow, cost of voting, and expected social image. Individuals vote if the net expected utility of doing so is positive:

$$pV + g - c + N\left[\max\left(s_V, s_N - L\right) - \max\left(s_N, s_V - L\right)\right] \geq 0. \tag{1}$$

The first three terms in expression (1) capture the standard model of voting. The first term is the expected utility of being pivotal (Downs, 1957), with a pivotality probability p and value V assigned to deciding the election. The second term, g, is the warm glow from voting (as in Riker and Ordeshook (1968)). The third term, $-c$, is the transaction cost of going to the polls. Since our experimental design does not focus on these components, only their sum will matter, which we denote by $\varepsilon = pV + g - c$. We assume $-\varepsilon$ has c.d.f. H.

The crux of the model is the fourth term, the social-image motivation to vote (in the spirit of Harbaugh (1996)). An individual expects to be asked whether she voted N times, and has to decide whether to be truthful or to lie. Assume first that she has voted. In this case, she can (truthfully) state that she voted, which earns her utility s_V; or she can lie and look like a non-voter, which earns her utility s_N minus a psychological lying cost L. Therefore, the utility a voter receives when being asked about her turnout is $z^v \equiv \max\left(s_V, s_N - L\right)$. Now assume that she did not vote. In this case, she can either state the truth and obtain the utility from appearing to be a non-voter, s_N, or lie and obtain s_V minus the lying cost L. Hence, the utility of being asked for a non-voter is $z^{nv} \equiv \max\left(s_N, s_V - L\right)$. The term in square brackets in (1) is therefore the net utility gain from voting due to being asked once.

The terms s_V and s_N capture how much the individual cares about being seen as a public good contributor (voter), or not, by others. These terms can be understood as reduced-form representations of a signalling model, such as Benabou and Tirole (2006) and Ali and Lin (2013). Experimental evidence suggests that information about whether a person votes affects how favorably they are viewed by others (Gerber et al. 2012).

The term L captures the utility cost of lying. We assume that the cost of lying is non-negative, $L \geq 0$, and additive with respect to the social-image term. The assumption of positive lying costs is motivated both by introspection and by experimental evidence documenting that in cheap talk communication games, which are similar to survey questions, a sizeable portion of subjects prefer to tell the truth even when lying is profitable.

In the general case, we do not impose any restrictions on s_V and s_N, but we consider two special cases: (i) *Pride in Voting* ($s_V > 0$): individuals care (positively) about stating that they are voters; (ii) *Stigma from Not Voting* ($s_N < 0$ and $s_V - L < 0$): individuals dislike both (truthfully) admitting to being non-voters and (untruthfully) saying that they are voters. Notice that both conditions could hold, for $s_V > 0 > s_N$, provided L is large enough.

Using the abbreviated notation ε for the other reasons to vote, we can rewrite the voting condition (1) as $N\Phi\left(s_V - s_N, L\right) + \varepsilon \geq 0$, where

$$\Phi\left(s_V - s_N, L\right) = \begin{cases} L & \text{if } s_V - s_N \geq L \\ s_V - s_N & \text{if } -L \leq s_V - s_N < L \\ -L & \text{if } s_V - s_N < -L. \end{cases} \tag{2}$$

As expression (2) shows, voting depends on the net social-image value $s_V - s_N$ and on the cost of lying L. Figure 1 displays $\Phi\left(s_V - s_N, L\right)$ as a function of $s_V - s_N$ for $L = 10$ and makes it clear that, in order for social image to contribute to voting, the net utility $s_V - s_N$ must be non-zero and the lying cost L must be positive. If either of these conditions is not met, then the individual either does not care about image, or can always signal the best-case scenario, irrespective of her true actions. Also notice that as long as individuals prefer to signal that they are voters ($s_V - s_N > 0$), the net value of being asked for voting is weakly positive.

Door-to-Door Survey. To estimate this model, we design a door-to-door survey in which individuals are asked, among other questions, whether they voted. We model the behavior of an individual whose home is visited by a surveyor. If the visit is pre-announced by a flyer and the person notices the flyer (which occurs with probability $r \in (0, 1]$), she can alter her probability of being at home and opening the door. A "survey flyer" (denoted by F) informs the reader when the surveyor will visit, but leaves the content of the survey unspecified. An "election flyer" (denoted by FE) additionally informs the reader that the survey will be about her voter participation in the previous election.

Once the surveyor visits the home, the respondent opens the door with probability h. If she did not notice the flyer (or did not receive one), h is equal to a baseline probability $h_0 \in (0, 1)$. If she noticed the flyer, she can optimally adjust the probability to $h \in [0, 1]$ at a cost $c\left(h\right)$, with $c(h_0) = 0$, $c'(h_0) = 0$, and $c''(\cdot) > 0$. That is, the marginal cost of small adjustments is small, but larger adjustments have an increasingly large cost. We allow for corner solutions at $h = 0$ or $h = 1$. In the estimation, we assume $c\left(h\right) = \left(h - h_0\right)^2 / 2\eta$.

If the individual is at home at the time of the surveyor's visit, she must decide whether to complete the survey. Consumers have a baseline utility s of completing a generic 10-minute survey for no monetary payment. The parameter s can be positive or negative to reflect that individuals may find surveys interesting, or they may dislike surveys. In addition, individuals receive utility from a payment m and disutility from the time cost c, for a total utility from survey completion of $s+m-c$. The time cost c equals τv_s, where τ is the duration of the survey in fraction of hours, and v_s is the value of one hour of time. In addition, as in DellaVigna, List and Malmendier (2012), the respondent pays a social pressure disutility cost $S \geq 0$ for refusing to do the survey when asked in person by the surveyor. There is no social pressure if the individual is not at home when the surveyor visits. We further assume that the respondent is aware of her own preferences and rationally anticipates her response to social pressure. In addition to the baseline utility $s+m-c$ of doing a survey, there is the additional utility from being asked about voting, z^v for voters and z^{nv} for non-voters, as defined above.

We also vary whether the survey content is announced to the respondent when she opens the door with two 'announcement' treatments, $a \in \{I, NI\}$. When informed that the survey will ask about her voter participation ($a = I$), an individual will consider the utility of being asked about voting, z^i, while deciding whether to complete the survey. If she is instead not informed at the door ($a = NI$), she will neglect z^i - provided she has not already seen an election flyer. This announcement treatment is in the spirit of the election flyer treatment, but by design can only affect survey completion, not the probability of answering the door.

Finally, in some treatment cells we provide an incentive for the respondents to say that they did not vote; the incentive is either in terms of time—an 8-minute shortening of the survey duration—or money—an extra \$5 for 1 more minute of questions. We denote by I the monetary value of the incentive. By incentivizing the respondent to say she did not vote, a voter is provided an incentive to lie, and will lie if $s_N^v - L^v + I \geq s_V^v$. In contrast, a non-voter is provided an incentive to tell the truth, and will do so if $s_N^{nv} + I \geq s_V^{nv} - L^{nv}$. By comparing the treatments with and without incentive I, we estimate the distribution of $s_V - s_N + L$ for voters and of $s_V - s_N - L$ for non-voters. Note that this treatment is unanticipated, and hence does not appear in the respondent's decision to answer the door or participate in the survey.

Solution. Conditional on answering the door, the respondent of type $i \in \{v, n\}$ agrees to the survey if $s^i + m - c^i + z^i \geq -S^i$ assuming the respondent knows that the survey is about election and if $s^i + m - c^i \geq -S^i$ otherwise. Working backwards, consider a respondent of type i who sees a survey flyer (which does not mention the election questions). The decision problem of staying at home (conditional on seeing a flyer) is $\max_{h\in[0,1]} h \max\left(s^i + m - c^i, -S^i\right) - (h - h_0)^2 / 2\eta^i$, leading to the solution $h^{i*} = \max\left[\min\left[h_0 + \eta^i \max\left(s^i + m - c^i, -S^i\right), 1\right], 0\right]$. An increase in pay m or a decrease in the time cost c will increase the probability of being at home and completing a survey. The parameter η^i determines the elasticity with respect to incentives of home presence. Alternatively, for a respondent who sees the election flyer the solution is given by $h^{i*} = \max\left[\min\left[h_0 + \eta^i \max\left(s^i + m - c^i + z^i, -S^i\right), 1\right], 0\right]$. If $z^i > 0$, the respondent will stay at home with a weakly higher probability with the election flyer, compared to the survey flyer, and vice versa if $z^i < 0$.

Finally, for both the survey flyer and the election flyer, there is a variant with an opt-out box which makes avoidance of the surveyor easier. In this condition, agents can costlessly reduce the probability of being at home to zero. Formally, $c(0) = 0$ and $c(h)$ is as above for $h > 0$.[6] The optimal probability of being at home h^* remains the same as without the opt-out option if there is no social pressure and, hence, no reason to opt out (since the respondent can costlessly refuse to do the survey) or if the agent expects to derive positive utility from completing the survey. In the presence of social pressure, however, the respondent opts out if the interaction with the surveyor lowers utility.

The following Propositions summarize the testable predictions about the impact of the election flyer (Propositions 1 and 2), about the incidence of lies about past turnout (Proposition 3) and about the expected number of times asked, which we manipulate in the get-out-the-vote intervention (Proposition 4).[7]

Proposition 1. (Pride in Voting) *With Pride in Voting, the probability of home presence $P(H)$ and of survey completion $P(SV)$ for voters is higher under the election flyer than under the survey flyer: $P(H)^v_{FE} \geq P(H)^v_F$ and $P(SV)^v_{FE} \geq P(SV)^v_F$. Parallel results hold for the opt-out flyers: $P(H)^v_{OOE} \geq P(H)^v_{OO}$ and $P(SV)^v_{OOE} \geq P(SV)^v_{OO}$ The probability of survey completion for voters is higher when informed at the door that the survey is about voting: $P(SV)^v_I \geq P(SV)^v_{NI}$.*

Proposition 2. (Stigma from Not Voting) *With Stigma from Not Voting, the probability of home presence $P(H)$ and of survey completion $P(SV)$ for non-voters is lower under the election flyer than under the survey flyer: $P(H)^{nv}_{FE} \leq P(H)^{nv}_F$* and *$P(SV)^{nv}_{FE} \leq P(SV)^{nv}_F$. Parallel results hold for the opt-out flyers: $P(H)^v_{OOE} \leq P(H)^v_{OO}$ and $P(SV)^v_{OOE} \leq P(SV)^v_{OO}$. The probability of survey completion for non-voters is lower when informed at the door that the survey is about voting: $P(SV)^{nv}_I \leq P(SV)^{nv}_{NI}$.*

Proposition 3. (Lying about Voting). *If the net social-image utility is positive, the probability of lying about past voting, $P(L)$, should be zero for voters and larger for non-voters assuming no incentives to lie ($I = 0$): $P(L)^v = 0 \leq P(L)^{nv}$ for $s_V - s_N > 0$. For any social-image utility, the probability of lying is (weakly) increasing in the incentive I for voters and (weakly) decreasing in I for non-voters: $\partial P(L)^v / \partial I \geq 0$* and *$\partial P(L)^{nv} / \partial I \leq 0$.*

Proposition 4. (Times Asked) *The probability of voting is increasing in the number of times asked N if the social-image utility is positive and lying costs are positive: $\partial P(V) / \partial N \geq 0$* for $s_V - s_N > 0$ and $L > 0$.

[6] This formalization allows a costless reduction of h to 0 but not to other levels. This is not a restriction because agents who prefer to lower h below h_0 (at a positive cost) will strictly prefer to lower h to 0 at no cost.

[7] The proofs are in the Appendix.

3 Experimental Design

Logistics and Sample. We employed 50 surveyors and many flyer distributors, mostly undergraduate students at the University of Chicago, who were paid $10.00 per hour. All surveyors conducted surveys within at least two treatments, and most over multiple weekends.[8] The distribution of flyers took place on Fridays and Saturdays, and the field experiment took place on Saturdays and Sundays between July 2011 and November 2011. The locations are towns around Chicago shown in Figure 2.[9] Each surveyor is assigned a list of typically 13 households per half-hour on a street (constituting a surveyor-route), for a daily workload of 8 routes (10am-12pm and 1-3pm). Every half-hour, the surveyor moves to a different street in the neighborhood and begins a new route of 13 homes, typically entering a different treatment in the next route. Surveyors do not know whether a treatment involves a flyer, though they can presumably learn that information from observing flyers on doors.

To determine the sample in each of the towns visited, we obtain voting records from the Election unit of the Cook County Clerk's office in January 2011. We begin with the full sample of addresses with at least one adult registered to vote. We then reduce the sample to households with homogeneous voting records in the congressional elections of November 2010: either every registered voter at the address voted in 2010, or no one did. Next, we randomize these households to a treatment at the surveyor-route level. Houses are grouped into surveyor-routes, which are then randomized to treatments. The treatment is a combination of four crossed interventions: (i) flyer treatments, (ii) payment and duration of the survey, (iii) survey content announcement at the door, and (iv) incentives to claim non-voter status.

Treatments. Each household was randomized into five flyer treatments with equal weights: *No Flyer*, *Survey Flyer*, *Election Flyer*, *Opt-Out Flyer*, and *Election Opt-Out Flyer*. Households in the *No Flyer* treatment receive no flyer. Households in the *Survey Flyer* treatment receive a flyer on the doorknob announcing that a surveyor would approach the home the next day within a specified hour (e.g., 3pm - 4pm, see top left example in Figure 3). Households in the *Election Flyer* treatment receive a similar flyer, with the added information that the survey will be about '*your voter participation in the 2010 congressional election*' (second flyer from left in Figure 3). Households in the *Opt-Out Flyer* treatment receive a flyer as in the *Survey Flyer* treatment, except for an added check-box which the household can mark if it does not wish to be disturbed (third flyer from left in Figure 3). Similarly, the flyer in the *Election Opt-Out Flyer* treatment has an added opt-out check box. The flyers were professionally produced.

A second crossed randomization involves the duration of the survey as well as the compensation offered (if any) for completing the survey. The bottom row of Figure 3 displays flyers for

[8] Additional details about the experiment, including the recruitment process, are in the Online Appendix.

[9] Arlington Heights, Elk Grove Village, Evanston, Glenview, Hoffman Estates, Lincolnwood, Mount Prospect, Northbrook, Oak Park, Park Ridge, Schaumburg, Skokie, Streamwood, Wilmette, and Winnetka. On almost all days, we visited one or two towns on a given day.

the three treatments: (5-Minutes, No Payment), (10-Minutes, $10 Payment), and (5-Minutes, $10 Payment), each sampled with equal probability. In each of these treatments we reiterated the compensation and duration at the door.

The third set of crossed treatments involves how the surveyors described the survey once, after a knock on the door, a household member answered. The respondents were told "*We are conducting confidential _ _ _ minute surveys in _ _ _ today. [You would be paid $ _ _ _ for your participation.]*", with the empty fields filled depending on the payment and duration treatments and the assigned town. The *No Information* group was then simply asked "*Do you think you might be interested?*". The *Information* group was instead told "*The survey is about your voter participation in the 2010 congressional election. Do you think you might be interested?*". Hence, the Information treatment provides information about the content of the survey in a similar way to the Election Flyer treatment. Respondents in the *Election Flyer* or *Election Opt-out Flyer* already knew about the content, provided they read the flyer. The top part of Figure 4 summarizes this first set of crossed treatments.

The fourth set of crossed treatments, summarized at the bottom of Figure 4, involves incentives to affect the response to a turnout question. In control surveys, individuals are simply asked whether they voted in the 2010 congressional election. For a subject in a 10-minute, $10 survey in the treatment group, we offer an 8-minute incentive to the respondent to state that he or she did not vote. After the first question in the survey, the surveyor reads aloud: '*We have 10 minutes of questions about your voter participation in the 2010 congressional election, but if you say that you did not vote then we only have 2 minutes of questions. Either way you answer you will be paid $10. That is, we have 10 minutes of questions, but if you tell us no to the question "did you vote in the 2010 congressional election " then we only have 2 minutes of questions to ask. Regardless of your answer you will earn $10.*' The surveyor then points to where the survey ends if the respondent answers 'no', in which case the survey is indeed much shorter.

For respondents assigned to a 5-minute survey, we did not assign a time discount which could only have been a modest 3-minute reduction. Instead, we provide a monetary incentive to the treatment group as follows (with the material in brackets applying only to the (5-Minutes, $10 Payment) conditions): '*We have 5 minutes of questions about your voter participation in the 2010 congressional election, but if you say that you did not vote then we have 1 extra minute of questions and we will pay you an extra $5 for answering these additional questions [.IF PAID: for a total of $15]. If you say that you voted then we will just ask you the original 5 minutes of questions. [.IF PAID: and pay you $10 as promised.] That is, we have 5 minutes of questions, but if tell us no to the question "did you vote in the 2010 congressional election" then we have 1 extra minute of questions and you will earn an additional $5 for answering these questions.'* Conditional on a 5-minute or a 10-minute survey, we determined the incentive or no-incentive treatment with equal weights.

Finally, we followed the promises made: we pay the individuals as promised, and we conducted a longer survey when the survey was advertised as lasting 10 minutes rather than 5 minutes. Further, in the treatments with a lying incentive, if the subject responded 'no' to the turnout question, the survey duration and payment were altered as promised.

Sample. We reached a total of 14,475 households. From this initial sample, we exclude 1,278 observations in which the households displayed a no-solicitor sign (in which case the surveyor did not contact the household) or the surveyor was not able to contact the household for other reasons (including, for example, a lack of access to the front door or a dog blocking the entrance). The final sample includes 13,197 households.

4 Reduced-Form Estimates

Answering the door and survey completion. We present graphical evidence in Figure 5 on the share of households answering the door and completing the survey as a function of the survey details, pooling across the five flyer treatments. Voters are very responsive to incentives, going from 33 percent answering the door for a \$0, 5-minute survey to 39 percent for the \$10, 5-minute survey. Hence, a \$10 incentive induces a 6 percentage point (20 percent) increase in the share answering the door. The effect is similarly large for the share completing the survey, a 6 percentage points (45 percent) increase. The elasticity of non-voters with respect to incentives is smaller with regards to answering the door, but is large with respect to survey completion: 5 percentage points (62 percent).

Having established that households are responsive to the survey incentives, we turn to the key flyer treatment—whether the flyer informs the household about the election question. Figure 6a plots the results for voters, pooling across the different survey durations and payment incentives. We do not observe much difference for voters in the share answering the door, or the share completing the survey, between the Survey Flyer and the Flyer Election treatments. In the Opt-out treatments, we observe a *decrease* in the share answering the door and in the share completing a survey when the survey informs about the election. Thus, there is no evidence of pride from voting, and it appears that voters prefer not to be asked whether they voted.

For non-voters (Figure 6b), the difference between the Flyer and the Flyer Election treatments is large: there is a 6 percentage point drop (20 percent) in the probability of answering the door. The size of this effect is comparable to the effect of a \$10 incentive to complete the survey. There is a similar 3 percentage point (25 percent) decrease in the share completing a survey when the flyer announces the election question. The impacts are consistent but smaller in the opt-out treatments, with a 1.5 percentage point (15 percent) decrease in the share answering the door when the flyer mentions elections. These results indicate strong avoidance of non-voters, pointing to shame from admitting to not voting and disutility from lying.

These findings may depend on the context. The results of the 2010 congressional elections were very disappointing for Democrats, including in Illinois the loss of President Obama's seat in the Senate, and correspondingly positive for Republicans. The lack of evidence for pride among voters may well be due to disappointment, given that the neighborhoods visited were largely Democratic. While our results are from a single election, we can differentiate the response based on the primary registration. In Figure 7 we present separate results for households with voters who participated in Republican primaries (left panel) versus households with voters registered in Democratic primaries (right panel).[10] Indeed, we detect sizeable sorting in by Republican voters in response to the election flyer, indicative of pride in voting in an election with positive results for the party. Among Democratic voters, instead, we observe sorting out as in the overall results, consistent with disappointment about the election. Among voters who did not participate in a primary (not shown), we also detect sorting out.

We now examine the effects of announcing the survey content at the door. Figure 8 plots survey completion rates by the door announcement type (Informed or Not Informed), pooling across all the flyer treatments. For voters, the effects of the door announcements are similar to those of the flyer announcements: there is no increase in survey completion from being informed about the voting question, and thus no evidence of pride. But non-voters also show essentially no effect on survey completion from being informed at the door. This is in contrast to the flyer treatments, where the election flyer leads to a sharp drop in answering the door and in survey completion by non-voters. We speculate that the difference (not captured in the model) could be that the flyer gives individuals time to think through the decision problem, while they must respond immediately when warned only at the door.

In Table 1, we present the regression analysis underlying Figures 5, 6a and 6b, 7, and 8 both with no controls and with fixed effects for surveyor i, day-town t, and hour-of-day h. We estimate, separately for voters and non-voters, the OLS regression:

$$y_{i,j,t,h} = \alpha + \Gamma T_{i,t,h} + \eta_i + \lambda_t + \zeta_h + \varepsilon_{i,j,t,h} \tag{3}$$

where the dependent variable $y_{i,j,t,h}$ is, alternatively, an indicator for whether individual j opened the door (y^H) or agreed to complete the survey (y^S). The vector $T_{i,t,h}$ contains indicators for the various survey treatments, with the baseline No-Flyer treatment for a \$0, 5 minute survey as the omitted group. We cluster the standard errors at the surveyor$\times$date level.[11] Table 1 shows that the results shown in the previous figures are robust to the inclusion of the surveyor, date-location, and hour fixed effects. In the Online Appendix we present two sets of

[10] We record the most recent participation in primary elections by any registered member of the household. We define as 'households with registered Republican voters' households where at least one voter has voted in a Republican primary, and no voter has voted in a Democratic primary. Vice versa for the definition of households with registered Democrats.

[11] For space reasons, the specification in Table 1 assumes an additive effect between the flyer treatments, the payment and duration treatments and the door information treatments. The empirical moments used for the estimation, listed in Appendix Table 1, are more disaggregated.

robustness results. In Online Appendix Table 1 we allow for different effects of a surveyor on different dates and location by including surveyor*date*location fixed effects; the results are unaffected. In Online Appendix Table 2 we present separate estimates for the first two month of the experiment (July and August 2011) and the next two months (October and November 2011); the results are comparable.[12] Finally, in Online Appendix Table 3 we present the results split by political registration, as in Figure 7.

Lying about voting. Next, we estimate the rates at which voters and non-voters misrepresent their voting behavior, and how these lies respond to the randomized incentives to lie (for voters) or to tell the truth (for non-voters). For the sample of individuals who completed the survey, we estimate the OLS regression

$$y_{i,j} = \alpha + \Gamma T_{i,j} + \eta_i + \varepsilon_{i,j} \tag{4}$$

where $y_{i,j} = 1$ if individual j lied about her voting behavior to surveyor i, and 0 otherwise, and $T_{i,j}$ is an indicator for whether respondent j is provided an incentive to say she did not vote. Due to the smaller sample, only surveyor fixed effects η_i are included in regressions.

In Table 2 and Figure 9, we present the results from these estimations. Recall that the incentive was always to say that one did not vote. Thus, we expect voters in the treatment condition to lie more than in the control, and non-voters to lie less. In Panel A of Table 2, to maximize power we pool across all survey treatments and across the 8-minute and $5 incentive. Note first that non-voters, in the absence of any lying incentive, lie about 46 percent of the time about past turnout. This rate is within the range of previous results using the American National Election Studies and validated voter records (Silver, Anderson, and Abramson, 1986), and indicates that non-voters care about the social image that they convey. We also observe a 12 percent lying rate for voters, which could be explained by measurement error in the match to the voting records, or by a genuine preference among some voters to look like a non-voter.[13]

Turning to the effect of the incentives, the treatments have a small effect on voters: they lie 2.7 percentage points more when incentivized to do so, which is not statistically significant at conventional levels. For non-voters, in contrast, the effect is a highly significant 12 percentage point (25 percent) decrease in lying rates. Thus, voters appear to greatly dislike lying and claiming to be non-voters (relative to telling the truth), while non-voters are more easily moved between telling the truth and falsely claiming to be voters.

Do the results differ for the 8-minute time discount versus the $5 incentive? Figure 9 shows that the results are very similar for the two types of incentives, especially for non-voters, suggesting an implied value of time of about $35 per hour. Panels B-D in Table 2 further show that the results are similar whether the 5-minute survey was paid or unpaid.

[12] We did not run the experiment in September 2011.

[13] Notice that non-registered voters do not appear in our voting records. Hence, some of the households which we classify as 'voting households' may include some non-voters, accounting for some of the lying rate for these households. In the Structural Estimates, we present results which allow for measurement error.

Summary. To summarize the reduced form results, among voters we find little sorting on average into the home in the election flyer treatment, and therefore little evidence of pride in voting on average (though there is evidence among Republicans). But this does not imply that social image does not motivate their voting behavior. In fact, even with substantial incentives of \$5 earned or 8 minutes saved, over 85% of voters refuse to say they did not vote. This indicates that voters have a high lying cost L^v, a low social-image value of being a non-voter s^v_N , or both. Both these factors induce a high social-image value of voting. For non-voters, we find substantial sorting out in the election flyer treatment, indicating that that non-voters experience stigma on average from not-voting. Further, close to half of non-voters lie and claim to be voters when asked. This implies that on average they are indifferent between the options: $s^{nv}_V - s^{nv}_N = L^{nv}$. A \$5 incentive reduces lying by 25%, indicating that a substantial share of non-voters are close to the margin in their decision to tell the truth or lie. In the next section, we utilize all the experimental treatments to estimate the social-image value of voting.

5 Structural Estimates

Set-up. To estimate the model of Section 2, we impose additional assumptions, some of which are relaxed below. Since all parameters are allowed to differ between voters and non-voters, for simplicity we omit the superscript $i = \{v, nv\}$. We assume that the social-image variables s_V and s_N are independently normally distributed across individuals, with differing means μ_V and μ_N but the same standard deviation, $\sigma_V = \sigma_N$, which we denote by σ_{SI}. The normality assumption allows for individuals who prefer the social image associated with not voting ($s_V < s_N$). We also assume a normal distribution with parameters μ_s and σ_s for the utility s of completing an unpaid 10-minute survey, as well as a quadratic cost of changing plans to be at home, $c(h) = (h - h_0)^2 / 2\eta$.

The key parameters of interest are: (i) μ_V, the mean social-image utility from saying that one voted; (ii) μ_N, the mean social-image utility of saying one did not vote; (iii) σ_{SI}, the standard deviation of the social-image utilities; (iv) L, the lying cost. In the benchmark specification, the parameters are identified up to the cost of lying; we thus display the results as a function of an assumed lying cost.

We also identify the following auxiliary parameters: (i) h_0, the baseline probability of opening the door; (ii) r, the probability of observing (and remembering) the flyer; (iii) η, the responsiveness of the probability of opening the door to the desirability of being at home; (iv) μ_s and σ_s, the mean and standard deviation of the baseline utility of doing a survey; (v) v^s, the value of one hour of time; (vi) S_s, the social pressure cost associated with saying no to the survey request. The total number of parameters is 11, including L, for voters and as many for non-voters for a total of 22 parameters.

To estimate the model, we use a classical minimum-distance estimator. Denote by $m(\xi)$

the vector of theoretical moments as a function of the parameters ξ, and by $\hat{m}$ the vector of observed moments. The minimum-distance estimator chooses the parameters $\hat{\xi}$ that minimize the distance $(m(\xi) - \hat{m})' W (m(\xi) - \hat{m})$. As a weighting matrix W, we use the diagonal of the inverse of the variance-covariance matrix. Hence, the estimator minimizes the sum of squared distances, weighted by the inverse variance of each moment. As a robustness check, we also use the identity matrix as the weight. We discuss further details in Appendix A.

To list the moments $m(\xi)$, we introduce the following indices: $i \in \{v, nv\}$ indicates voters and non-voters, $k \in \{NF, F, FE, OO, OOE\}$ indicates the flyer treatments, m indexes the payment and duration treatments, $m \in \{\$0, 5min; \$10, 10min; \$10, 5min\}$, a indicates the treatments on survey information at the door, $a \in \{I, NI\}$, and l indexes incentives to lie, $l \in \{NoInc, 8\,\mathrm{min}, \$5\}$. The moments $m(\xi)$ are: (i) the probability opening the door in survey treatments k, m, $P(H)^i_{k,m}$; (ii) the probability of completing the survey in survey treatments k, m, $P(SV)^i_{k,m}$; (iii) the probability of checking the opt-out box in the Opt-Out treatments, $P(OO)^i_{k,m}$ for $k \in \{OO, OOE\}$ (iv) the probability of completing the survey in the survey content treatments, given the flyer treatments: $P(H)^i_{a,k}$ and (v) the probability of lying about past turnout conditional on completing the survey, given incentive l, $P(L)^i_l$.[14] The empirical moments $\hat{m}$, 100 in total, are estimated in a first stage model using the same controls as in the main regressions, and are listed in Appendix Table 1.

Benchmark Estimates. The benchmark estimates (Table 3) provide no evidence of pride for voters: voters on average dislike informing others that they voted: $\mu^v_V = -5.86$ (se 1.94). However, voters dislike lying even more: $\mu^v_N - L^v = -24.81$ (se 5.14) is the disutility from saying that they did not vote. Notice that we cannot parse the extent to which this disutility is due to a large net social-image utility $\mu^v_V - \mu^v_N$ or a large lying cost L^v. There is substantial heterogeneity in these signalling values: $\sigma^v_{SI} = 12.35$ (se 3.10), implying that 32 percent of voters *do* take pride in saying they voted.

For non-voters, we estimate significant stigma on average from admitting that they did not vote: $\mu^{nv}_N = -4.61$ (se 2.40). On average, non-voters are nearly indifferent between admitting they did not vote and lying and claiming they voted ($\mu^{nv}_V - L = -4.23$, se 2.20), consistent with the finding that about half of non-voters lie in the control treatments. Heterogeneity across individuals is sizeable but smaller than for voters: $\sigma^{nv}_{SI} = 6.20$ (se 1.29).

Turning to the auxiliary parameters, we estimate that on average neither voters nor non-voters like unpaid surveys, but there is a substantial heterogeneity, with voters being more likely to complete unpaid surveys. (Voters are likely public good providers generally). The

[14] We present pooled moments across some of the treatments for two reasons. In some cases we do not expect any impact of the treatment on the relevant moment, such as of the lying incentives on the probability of opening the door or completing the survey. In other cases, we pool to keep the list of moments readable and to guarantee a sizeable sample in each cell, when the model does not imply important differences across the pooled treatments; for example, we do not consider the impact of the survey content treatment separately as a function of the survey duration and payment.

estimated time value is \$65 per hour for voters and \$19 for non-voters, a difference consistent with the strong positive correlation between income and turnout. Voters are also estimated to incur higher social pressure costs from declining to participate in the survey ($S_s^v = \$1.76$ versus $S_s^n = \$0.06$) and a lower elasticity of home presence ($\eta^v = 0.13$ versus $\eta^n = 2.86$), although the elasticity for non-voters is imprecisely estimated.[15]

Value of Voting. Using the estimates, we compute the average social-image value of voting due to being asked *once*, $\int \Phi(s_V - s_N, L)\, dF(s_V, s_N)$. Since the benchmark model is identified up to the lying cost, we cannot point identify this value. We can, however, plot the social-image value of voting as *a function of* the lying cost L for a range of plausible values, shown in Figure 10a for both voters and non-voters. If lying is entirely costless ($L = 0$), the social-image value is zero, since non-voters and voters can costlessly claim their preferred social image. As the underlying cost of lying increases, the value of voting is inverse-U-shaped for voters, while it rises monotonically for non-voters. The intuition for voters is as follows. As L rises, as Figure 1 illustrates, the value of voting (weakly) increases in L for a *given* positive $s_V - s_N$, since lying becomes costlier. However, as L increases, the net estimated social image $s_V - s_N$ mechanically declines. (The data pins down the value for voters of saying that they did not vote, $s_N - L$; as L increases, s_N must increase to compensate, lowering $s_V - s_N$) Initially, the first force dominates since the lying cost is likely to be binding. For high enough lying cost, however, the second force shifting the net social image dominates, ultimately leading to a negative value of voting for high enough L. The intuition for non-voters is parallel.

The social-image value of voting in Figure 10a has two important implications for voting. First, even if we do not know the lying cost, for a range of plausible values the correspondent value of voting is quite flat. For a lying cost in the range between \$2 and \$15, the value of voting due to being asked once for voters lies between \$1.5 and \$4. To provide a benchmark estimate from a different context, In Appendix E we estimate the lying cost using the data from Erat and Gneezy (2012), a representative cheap talk laboratory experiment, and obtain an estimate for the lying cost in this range, of \$7. Hence, even though we are not point identified, the range of uncertainty for the value of voting is not too large provided one agrees on the range for L. (For non-voters instead the estimates of the value of voting increase sharply with L). Second, we can contrast the value of voting for voters and non-voters, provided we assume the same lying cost. For values of the lying cost up to \$7, the estimated value of voting is larger for voters than non-voters, consistent with observed cross-sectional differences.

To obtain the ultimate value of voting because others ask, we scale up the value of being asked once by the expected number of times asked, N. We measure this parameter with survey questions on how often the survey respondents has been asked whether they voted in the 2010 congressional election by friends, relative, coworkers, and other people. The Online Appendix Figure 1 displays the c.d.f. of the total number of times asked: 60 percent of respondents

[15]We discuss the imprecision of the estimates of η and S in Appendix C.

report being asked at least once, and 20 percent report being asked more than 10 times. On average, hence, respondents report being asked around 5 times for the 2010 congressional election, with similar magnitudes for voters and non-voters ($N^v = 4.89$ and $N^{nv} = 5.33$). The Figure also reports the number of times people report being asked for the 2008 presidential election: the average is about twice as high, with 40 percent of people reporting to be asked at least 10 times. This number is consistent with the corresponding figures in the Cooperative Congressional Election Study as reported in Gerber et al. (2012).

To obtain the total value of voting *because others ask* in the 2010 election, we multiply the value of voting due to being asked once by the average times asked, leading to an overall value of voting for voters, for the range of lying costs, between \$6.5 and \$15, sizeable magnitudes (Figure 10b and Table 3).[16] We can also use these estimates to conjecture the social-image value of voting for presidential elections. Assuming that the social-image parameters for presidential elections are at least comparable to the ones in congressional elections, we can multiply the value of voting due to being asked once by the number of times people report being asked in the 2008 presidential election. The implied value of voting in presidential elections, also plotted in Figure 10b, is about twice as large, in the range of \$13-\$30. The model is therefore also consistent with the observed higher turnout in presidential elections.

Finally, we turn to the welfare effect of being asked once if one has voted (Table 3). For voters, this is the average value of $z^v = \max\left(s_V^v, s_N^v - L^v\right)$, and is estimated to be -4.63 (se 1.98). Interestingly, non-voters are estimated to have a less negative utility from being asked, -0.92 (se 2.15) because non-voters' social-image utility distribution is closer to zero.

Identification and Sensitivity Analysis. We next discuss the intuition for the main sources of identification, and provide supporting evidence by calculating the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). In particular, we compute a sensitivity matrix S, each element S_{ij} of which is the derivative of the estimated parameter $\widehat{\xi}_i$ with respect to the moment m_j.[17] We further normalize the measure so each element can be interpreted as the effect of a one standard deviation increase in the moment m_j on the expected value of the estimated parameter $\widehat{\xi}_i$, *holding other moments fixed* and measured in units of the asymptotic standard deviation of $\widehat{\xi}_i$. This calculation reveals the local sensitivity of the parameter estimates to each of the individual moments, and thus allows us to provide evidence on which features of the data and the experiment (locally) identify the model. Alternatively, the measure can be interpreted as revealing the sensitivity of the estimates to

[16] We estimate the social-image utility when asked by a surveyor. If social-image concerns or lying costs when interacting with friends, family and colleagues are higher than those in a one-shot interaction with a surveyor, then our estimates are likely to be lower bounds of the social-image value of voting.

[17] We first compute the absolute sensitivity matrix $S = (\hat{G}'W\hat{G})^{-1}\hat{G}'W$, where W is the weighting matrix used in the minimum distance estimation, and $\hat{G} \equiv N^{-1}\sum_{i=1}^{N}\nabla_{\xi}m_i(\widehat{\xi})$ is the Jacobian of the moments with respect to the parameters. We then normalize the sensitivity measure as $\widetilde{S}_{ij} = S_{ij}\sqrt{\mathrm{Var}\left(m_j\right)/\mathrm{Var}\left(\widehat{\xi}_i\right)}$.

model misspecification.

First, consider the key social-image parameters, μ_V, μ_N and σ_{SI}. The difference in home presence and survey completion between the Flyer and Flyer-election, and between the Opt-Out and Opt-Out election treatments, play an important role. For voters, they help identify the mean social-image utility μ_V^v. For non-voters, given that on average half of non-voters lie in our sample (absent incentives to do otherwise), the average social-image utility from admitting to not voting μ_N^n must approximately equal the utility from lying, $\mu_V^n - L^n$. Hence, the sorting response for non-voters to the election surveys identifies both μ_N^n and $\mu_V^n - L^n$. A similar role is played by the difference in survey completion between the Information and No Information treatments. Finally, the response to the lying incentives is crucial for identifying the heterogeneity in social image σ_{SI} and the average utility difference between answering truthfully and lying. For example, an 8 minute incentive reduces the share of non-voters lying by 12 percentage points (Table 2, Panel D), implying $\Pr\left(s_N^{nv} < s_V^{nv} - L^{nv} < s_N^{nv} + (8/60)\, v_s^{nv}\right) = 0.12$ or $\Pr\left(0 < s_V^{nv} - L - s_N^{nv} < (8/60)\, v_s\right) = 0.12$, where $s_V^{nv} - L - s_N^{nv}$ is normally distributed with variance $2\sigma_{SI}^{nv2}$.

The parameter sensitivity calculations at the estimated values largely confirm these intuitions. Appendix Figure 1a plots the normalized sensitivity measures for voters for μ_V^v, μ_N^v and σ_{SI}^v as a function of each of the 50 empirical moments for voters. The home presence (PH) and survey completion (PSV) moments for the Flyer Election (WE) and Opt-Out Election (OOE) treatments largely show the expected effect: an increase in the moment (holding fixed the corresponding moments in the Flyer and Opt-Out treatments) leads to an increase in the estimated μ_V^v and μ_N^v values.[18] The opposite pattern obtains for the Flyer and Opt-Out moments (holding constant the Flyer Election and Opt-Out Election moments), since it is the difference between the two that is most relevant. As predicted, an increase in survey completion in the Information treatment relative to the No Information treatment also increases μ_V^v and μ_N^v, since it implies a larger relative attractiveness of answering a survey about voter turnout. Finally, the estimated heterogeneity in signaling values σ_{SI}^v responds strongly to the lying incentive moments. More lying by voters in the incentive-to-lie treatments implies a greater mass of individuals close to the margin, and thus a lower estimate of σ_{SI}^v (i.e. a negative sensitivity value). More lying by voters when incentivized also implies a higher social-image value of not voting μ_N^v relative to voting (μ_V^v).

Appendix Figure 1b plots the parallel analysis for non-voters. Similarly to the case of voters, the lying incentive moments play an important role, in the expected direction. Recall that non-

[18] An exception occurs in each case for the unpaid 5 minute survey, where a ceteris paribus increase in the WE moments generally *reduces* the mean social image parameters. The fact that this decrease occurs not just for WE but also for W and the PSV_NW moments suggests that is through a mechanism unrelated to the one described above, and involves instead other important parameters which feed back to the estimation of μ_V and μ_N.

voters are incentivized to tell the truth. Thus, a ceteris paribus increase in lying in the incentive treatment implies a smaller response to the incentive, and thus fewer marginal respondents due to a higher σ_{SI}^n. Conversely, an increase in lying in the control group implies a larger response to incentive, and thus a lower σ_{SI}^n. Higher lying in the control additionally implies a higher value of μ_V^n relative to μ_N^n. Compared to to the case of voters, the Information treatments play a more important role for identification of μ_V and μ_N for non-voters. As we discussed above, the Information treatments do not provide evidence of sorting out, in contrast to the strong sorting out patterns in both the Flyer Election and the Opt-Out Election treatments. At the estimated parameters, since we explain the response to the Information treatments poorly for non-voters, the estimates become quite sensitive to these moments. This motivates a robustness check below in which we do not utilize these moments.

As for the auxiliary parameters, the mean and standard deviation of the value of completing a survey, μ_s and σ_s, are identified from the survey completion rates for different monetary incentives (Online Appendix Figures 2a and 2b). The value of time v_s is identified from the comparison between payment increases (from \$0 to \$10) and duration decreases (from 10 to 5 minutes), and partly also by the response to the 8 minute time saving offered in the lying incentive (Online Appendix Figure 3). The baseline probability of answering the door, h_0, is pinned down by the share opening the door in the no-flyer treatments, and less directly by the share opting out in the opt-out treatments, since respondents are predicted to opt out only if they expect to be home in the first place (Online Appendix Figure 4). The probability of observing and remembering the flyer, r, is mainly identified by the fraction of households checking the opt-out box in the Opt-out treatment (10 to 13 percent), which equals $rh_0F_s(c-m)$, and by the fraction opening the door in these treatments (Online Appendix Figure 5). The elasticity of opening the door η with respect to incentives, and the social pressure S_s, are related to the share opening the door in the different survey treatments.[19] Identifying them separately is not obvious, since they often appear in the model in the product ηS. Indeed, Online Appendix Figure 6 shows that the two parameters respond in opposite directions to each moment, and also are identified by a wider range of moments. The difficulty in separately estimating η and S motivates the robustness checks below where we fix η.

Decomposing the Benchmark Estimates. To further highlight the identification, in Table 4 we re-estimate the model using subsets of the moments. When we use only the lying moments (Column 2), the levels of the social-image parameters μ_V and μ_N are not identified, but we can estimate the average *difference* $\mu_V - \mu_N$ (given a lying cost), as well as the heterogeneity σ_{SI}. Hence, holding μ_V fixed, we can estimate the other social-image parameters. The implied value of voting is then essentially the same as when using the much

[19] Consider a respondent of type i who dislikes answering a survey and hence will say no and incur the social pressure cost S_s. In the flyer treatment F, she will choose to be at home with probability $h_0^i - \eta^i r^i S_s^i$ (barring corner solutions for h).

richer set of moments. This highlights the key role that the lying treatments play in estimating the value of voting.

Alternatively, we utilize all the moments other than the experimental variation in incentives to lie (Columns 3 and 4). (We include the lying rates in the control group) These moments suffice to estimate all of the ancillary parameters as well as μ_V for voters. However, the other parameters are not identified because the heterogeneity term σ_{SI} is unidentified. Thus, we present two special cases with fixed low heterogeneity (Column 3, $\sigma_{SI}^{nv} = \sigma_{SI}^{v} = \$5$) and fixed high heterogeneity (Column 4, $\sigma_{SI}^{nv} = \sigma_{SI}^{v} = \$20$). The estimates reveal substantial sensitivity to the assumed σ. The key contribution of these sorting moments is not to the value of voting, but to the levels of the social-image terms and thus the welfare effect of being asked, which is not identified using only the lying moments.

Finally, in Column 5 we report estimates excluding the moments split by whether households are informed at the door about the election topic. When we exclude these moments, the estimated social-image parameters indicate a larger dislike of being asked about voting, especially for non-voters. This change, however, has essentially no effect on the estimated value of voting which, as we showed, largely depends on the lying incentives.

Robustness. In Table 5, we explore the robustness of the parameter estimates to alternative assumptions. First, we consider an alternative explanation of the results: the sorting out of non-voters may be due to a dislike of talking about politics, rather than any stigma from admitting to not voting. We thus allow for a utility of talking about politics which is independent of whether one voted or not (Column 2). With this extra parameter, we lose the ability to estimate a social-image parameter, but the estimated value of voting, which is identified by the lying treatments (provided a value of time), is unchanged.

Next, we consider two forms of measurement error. First, notice that the voting records do not include information about non-registered adults in a household. Since these individuals are necessarily non-voters, the person answering the door in an apparent voting household may actually be a non-voter. (This would explain why 10% of voting households appear to lie about voting even absent incentives to lie). In Column 3, we assume that 10% of respondents in voting households are actually non-voters. In Column 4, we allow for measurement error for both groups of households, and assume that 10% of respondents in a voting (or respectively, non-voting) household are non-voters (respectively, voters). Both specifications lead to a somewhat larger value of voting for a given lying cost.

We also consider a number of other robustness checks in Appendix Table 2. We show that the results are very similar if (i) we equally weight the moments instead of utilizing the inverse of the variances of each moment as weights; (ii) we assume positively correlated s_V and s_N, with correlation coefficient $\rho = 0.5$; (iii) we assume negatively correlated s_V and s_N, with $\rho = -0.5$. We can also, instead of estimating the elasticity of sorting η, fix it at either a large value ($\eta^v = \eta^{nv} = 1.0$) or a small value ($\eta^v = \eta^{nv} = 0.02$); these specifications are associated with a substantially worse fit, with some shifts in the social-image parameters.

Full Estimation. Thus far, we have been unable to identify the psychological cost of lying L separately from the utility terms s_V and s_N. However, we would be able to estimate all parameters, including the lying cost, if we made the additional assumption that voters and non-voters share the same social-image parameters. To see this, consider that the estimates allow us to identify μ_V^v and $\mu_N^v - L^v$ for voters, and μ_N^{nv} and $\mu_V^{nv} - L^{nv}$; if one assumes $\mu_V^v = \mu_V^{nv}$, $\mu_N^v = \mu_N^{nv}$ and $L^v = L^{nv}$, the value of the lying cost is identified.

In general, the assumption of equality for voters and non-voters is unpalatable. Different social-image parameters might be the very difference between voters and non-voters. The assumption, however, is less problematic in a subsample of voters and non-voters with similar voting histories. We eliminate always-voters and never-voters and consider households with individuals who vote some of the time, some of whom happened to vote in 2010, while others happened not to. Within this sub-group, individuals who voted in 2010 and individuals who did not vote in 2010 are more likely to be similar.

Formally, we use the individual-level voting records for all elections from 2004 to 2010, primaries included, to predict the probability that an individual will vote in November of 2010. We then restrict the sample to households where *all* registered individuals have a predicted probability of voting between 25 and 75 percent, leaving us with a sample of 5,901 households. Column 1 of Table 6 reports the estimation results. The mean utility of truthfully saying one voted (μ_V) is estimated near zero, with a net signalling utility $\mu_V - \mu_N$ of 4.50. The estimated lying cost is $L = \$4.63$ (se \$1.08), in the ballpark of plausible values. The implied total signaling value of voting in the 2010 elections is estimated at \$8.28 for voters and \$8.80 for non-voters (since the average number of times asked is slightly higher for non-voters).

The estimates in Column 1 require that not only the social signaling and lying parameters be the same across voters and non-voters, but that the auxiliary parameters be the same as well. In Column 2, we remove the latter assumption. Allowing for differences in the auxiliary parameters has very little effect on the key parameter estimates, but improves the fit of the model quite a bit, from an SSE of 132.49 (Column 1) to an SSE of 100.79. Finally, in Column 3 we report estimates where we allow all parameters to differ between voters and non-voters. This is the parallel of the benchmark specification in Table 3, and hence does not allow for point identification of all parameters, but is restricted to this special subsample. Allowing for this extra difference leads to only a relatively small increase in the fit, to an SSE of 97.20.

For the preferred specification in Column 2, we again calculate the Gentzkow-Shapiro sensitivity measures, but this time including the estimated lying cost parameter L. Appendix Table 7 shows that L is locally identified mainly by the lying incentives. A higher lying rate for voters when incentivized to lie results in a smaller estimated lying cost. Conversely, higher lying by non-voters when unincentivized implies a smaller lying cost. Greater survey completion in the unpaid election surveys (WE and OOE) also play a role, though with smaller effects. The sensitivity of the social-image parameters μ_V, μ_V and σ^v_{SI} and of the auxiliary parameters is similar to that reported for the benchmark estimates.

6 Get-out-the-vote Experiment

The experiments described above are designed to measure the value of voting without affecting voting itself. Yet, the model suggests a natural treatment to increase voter turnout. As Proposition 4 states, individuals with social-image motives are more likely to vote the more frequently they expect to be asked about voting, an expectation which we can manipulate experimentally.

In November of 2010 and of 2012, we set out to do just that in the suburbs of Chicago. In the five days before the election date, we posted a flyer on the doorknob of households in the treatment group informing them that '*researchers will contact you within three weeks of the election [...] to conduct a survey on your voter participation*'. Figure 11 shows the flyer for the 2012 election. (After the election, we follow up with a door-to-door visit, as advertised). Since this flyer could also impact turnout through a reminder effect, we compare this group to a group which received a flyer with a mere reminder of the upcoming election, also displayed in Figure 11. Finally, a control group received no flyer. After the election, we obtain the voting record for all individuals residing at the addresses targeted in this experiment, and we examine the impact on voter turnout.

Table 7 reports the results for both the November 2010 and the November 2012 intervention using an OLS specification: the dependent variable is an indicator for whether the individual voted in the specific election. Note that there may be multiple individuals at one address, each of which is a separate observation. The November 2010 experiment has a sample size of 31,306 individuals targeted. The turnout in the control group (which received no flyers) is 60.0 percentage points. Compared to this control group, the mere reminder had no effect, leading to an estimated decrease of .2 percentage point. Compared to the flyer with a mere reminder, the flyer with announcement of future question about voting raises turnout by 1.4 percentage points, a sizeable effect, albeit statistically insignificant. In Column 2, we add controls for the full history of voting of the households in all elections between 2004 and the election in question. Adding controls in a randomized experiment should not affect the point estimates if the experiment is conducted properly, but can reduce the residual variance, and hence increase precision. Indeed, the controls have very little impact on the point estimates, but they nearly halve the standard errors since past voting is highly predictive of future voting (the R^2 increases from 0.00 to 0.40). In this specification, the estimated effect of the flyer with announcement of future asking is an extra 1.3 percentage points in turnout, with a two-sided p-value of 0.12 (one-sided p-value of 0.06). While not quite statistically significant, the sizeable effect is certainly consistent with the predictions of the model.

Columns 3 and 4 display the estimates for the November 2012 election. In this later election, we were able to deploy a larger flyering team, guaranteeing a sample size of 93,805 individuals. Given the different nature of the election (presidential versus congressional), the baseline turnout in the control group is higher, at 73.1 percentage points. We find suggestive evidence that the reminder flyer itself may have increased turnout, with little evidence of a differential effect of the flyer with announcement of the future visit. In the specification with controls (Column 4), the differential effect is estimated to be 0.1 percentage points, not significant. The smaller effect in this second election is consistent with the fact that in this higher-stake election our intervention is competing with a high number of campaign materials; in addition, the higher baseline turnout leaves a smaller share of non-voters to be potentially convinced.

An important question is whether these estimated effects are consistent with the estimated value of voting in Section 5, or whether the model in Section 2 would imply a much larger get-out-the-vote effect. Using equation (1) for the turnout decision, we compute the implied increase in turnout due to an increase in N, the number of times asked about voting. We take the benchmark estimates for μ_v, μ_N, and σ_V, as well as N, in Table 3 (as a function of an assumed L). To obtain a predicted value of voting, we need in addition the mean μ_ε and s.d. σ_ε of the distribution of the other reasons to vote, ε. We pin down μ_ε using the observed turnout by group in 2010, and we consider two cases for σ_ε, a low case and a high case. For the low case, we assume that the standard deviation σ_ε of all the other reasons to vote is equal to the standard deviation of $s_V - s_N$, the net signalling utility. For the high case, we assume that the standard deviation σ_ε is ten times as large as the standard deviation of $s_V - s_N$. Since the implied get-out-the-vote effect is (approximately) decreasing in σ_ε, the two cases provide upper and lower bounds (Additional details are in Appendix D.)

In Figure 12, we plot the implied get-out-the vote effect as a function of L and for the two assumptions on σ_ε as the black lines. For the low calibration of σ_ε (continuous line), the implied get-out-the-vote effect increases with the assumed lying cost L and equals the estimated effect for a lying cost L of \$2. Even for sizeable larger lying costs ($L = 5$), the estimated get-out-the-vote effect, while larger than the estimated one at 2 percentage points, remains within the confidence interval of the estimates. Under the high calibration for σ_ε (dotted line), the implied get-out-the-vote effects are much smaller, in the order of 0.1 to 0.2 percentage points, not surprisingly since a higher variance of other reasons to vote reduces the impact of the additional social incentive to vote. This lower figure is still within the confidence interval of the estimates.

We also provide the implied calibration for presidential elections, though this calibration is admittedly tentative given that we do not estimate the parameters for the 2012 election. Still, we can take the 2010 parameters and assume, as in the data, a higher number of times asked and a higher turnout. As the red lines in Figure 12 show, the implied effect for presidential effects is proportionately smaller, in line with the smaller get-out-the-vote effects for 2012. While we should stress again that these implied magnitudes are tentative given the needed assumptions about the error term and the lying cost, they show that there is no disconnect between the estimates of the value of voting 'because others will ask' and the get-out-the-vote results.

The results are also consistent with the contemporaneous and independent results of Rogers and Ternovsky (2013) who similarly inform a treatment group that they may be called after the election about their voting behavior, and find a similarly positive impact on turnout. Importantly, it is intuitive that the effect sizes are much smaller than those found in Gerber, Green and Larimer (2008). Their intervention is conducted in a non-competitive primary election, and explicitly threatens to truthfully reveal one's voting record to the entire neighborhood, while also providing information about social norms (other's voting records) and a strong civic duty exhortation. In contrast, we operate in more competitive elections, do not provide civic duty messaging or information on neighbors, and warn of at most one additional question about voting, from a researcher who can be avoided or lied to.

7 Conclusion

We have presented evidence from a natural field experiment designed to estimate a social-image model of voting: individuals vote because they expect to be asked, and they anticipate the disutility associated with admitting to not voting, or with lying about voting. The results provide significant support for the model, and document substantial shame from admitting to not voting and little pride from conversely claiming to vote. The combination of three crossed experimental arms allows us to estimate the key social image parameters. The final piece of the puzzle – the lying cost – can be identified for a subsample of individuals with a medium propensity to vote. For this subsample, we obtain an estimate of the value of voting simply due to being asked of $8, a sizeable magnitude for a congressional election. These constitute among the first estimates of the private value of voting.

A methodological ingredient of this paper is the tight link between a simple model and the experimental design. This allows us not only to derive reduced-form results, but to use such results to estimate the underlying parameters using credible experimental variation and transparent identification. As such, this paper attempts to bridge a gap between two thriving, but largely separate literatures: the theoretical literature on voting and on social image, and the reduced-form field experiments on get-out-the-vote and turnout. We hope that methodologies similar to the ones in this paper will be useful in providing further insights.

References

[1] Allcott, Hunt. (2011). "Social Norms and Energy Conservation." *Journal of Public Economics,* 95 (9-10), 1082-1095.

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

[3] Andreoni, James, B. Douglas Bernheim. "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects" *Econometrica,* 77 (5), 1607–1636.

[4] Ariely, Dan, Anat Bracha and Stephan Meier. (2009). "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially" *American Economic Review*, 99(1), 544-555.

[5] Ashraf, Nava, Oriana Bandiera and Kelsey Jack. (2014) "No Margin, No mission? A field experiment on incentives for pro-social tasks" *Journal of Public Economics*, 120: 1-17.

[6] Benabou, Roland and Tirole, Jean, (2006). "Incentives and Prosocial Behavior," *American Economic Review,* 96(5), 1652-1678.

[7] Blais, Andre, (2000). *To Vote or Not to Vote: The Merits and Limits of Rational Choice Theory*, Pittsburgh: University of Pittsburgh Press.

[8] Leonardo Bursztyn, Michael Callen, Bruno Ferman, Ali Hasanain, and Noam Yuchtman. (2014). "A Revealed Preference Approach to the Elicitation of Political Attitudes: Experimental Evidence on Anti-Americanism in Pakistan" Working paper.

[9] Coate, Stephen and Michael Conlin. (2004). "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence" *American Economic Review*, 94(5), 1476-1504.

[10] Coate, Stephen, Conlin, Michael, and Moro, Andrea. (2008). "The Performance of Pivotal-Voter Models in Small-Scale Elections: Evidence from Texas Liquor Referenda," *Journal of Public Economics*, 92: 582-596.

[11] Conlin, Michael, O'Donoghue, Ted, and Vogelsang, Timothy J., (2007). "Projection Bias in Catalog Orders," *American Economic Review*, 97: 1217-1249.

[12] DellaVigna, Stefano, List, John. A., & Malmendier, Ulrike. (2012). "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics*, 127(1), 1-56.

[13] Downs, Anthony, (1957). *An Economic Theory of Democracy*, New York, Harper.

[14] Erat, Sanjiv and Gneezy, Uri. (2012). "White Lies," *Management Science*, 58(4): 723-733.

[15] Feddersen, Timothy and Alvaro Sandroni, (2006), "A Theory of Participation in Elections," *American Economic Review*, 96(4), 1271-1282.

[16] Finan, Frederico and Laura Schechter. "Vote-Buying and Reciprocity", *Econometrica*, 80(2): 863-882.

[17] Gentzkow, Matthew and Jesse Shapiro. (2013) "Measuring the sensitivity of parameter estimates to sample statistics." Unpublished Manuscript.

[18] Gerber, Alan S., and Donald P. Green. (2000). "The Effects of Canvassing, Direct Mail, and Telephone Contact on Voter Turnout: A Field Experiment." *American Political Science Review* 94: 653-63.

[19] Gerber, Alan S., Green, Donald P., and Larimer, Christopher W., (2008). "Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment," *American Political Science Review*, 102(1): 33-48.

[20] Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. (2012.) "Social Judgments and Political Participation: Estimating the Consequences of Social Rewards and Sanctions for Voting."

[21] Green, Donald P. and Alan S. Gerber. (2008). *Get Out the Vote: How to Increase Voter Turnout,* Brookings Press.

[22] Harbaugh, William T., (1996). "If People Vote Because They Like to, Then Why Do So Many of Them Lie?" *Public Choice*, 89(1-2): 63-76.

[23] Harsanyi, John, (1977). "Morality and the Theory of Rational Behavior," *Social Research*, 44(4): 623-56.

[24] Laibson, David, Andrea Repetto, and Jeremy Tobacman, (2007). "Estimating Discount Functions with Consumption Choices over the Lifecycle," Working paper.

[25] Lacetera, Nicola, and Mario Macis. (2010). "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme" *Journal of Economic Behavior & Organization,* 76 225–237.

[26] Ledyard, John O., (1984). "The Pure Theory of Large Two-Candidate Elections," *Public Choice*, 44: 7–41.

[27] Palfrey, Thomas and Rosenthal, Howard, (1983). "A Strategic Calculus of Voting," *Public Choice*, 41: 7–53.

[28] Palfrey, Thomas and Rosenthal, Howard, (1985). "Voter Participation and Strategic Uncertainty," *American Political Science Review*, 79: 62–78.

[29] Passarelli, Francesco and Guido Tabellini. (2013) "Emotions and Political Unrest" Working paper.

[30] Perez-Truglia, Ricardo and Guillermo Cruces. (2013) "Social Incentives in Contributions: Field Experiment Evidence from the 2012 U.S. Presidential Campaigns", unpublished manuscript.

[31] Powell, M.J.D., "Variable Metric Methods for Constrained Optimization," *Mathematical Programming: The State of the Art*, (A. Bachem, M. Grotschel and B. Korte, eds.) Springer Verlag, pp 288-311, 1983.

[32] Riker, W.H. and Ordeshook, P.C. (1968). "A theory of the calculus of voting." *American Political Science Review* 62: 25-42.

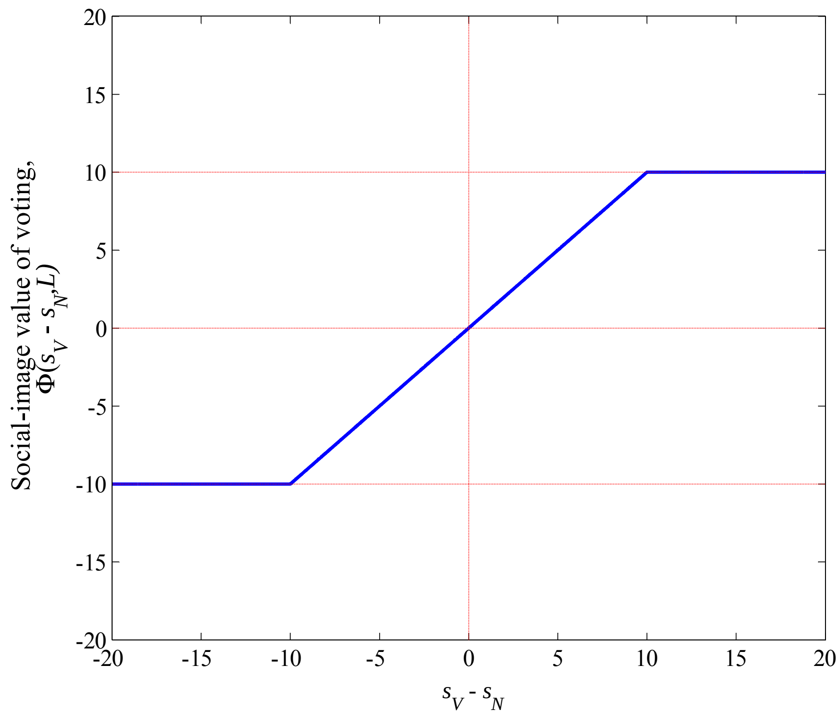
[33] Rogers, Todd and John Ternovski. (2013). "We May Call Post-Election to Ask if You Voted: Accountability and a Behavior's Importance to the Self." Working paper.

[34] Shue, Kelly and Erzo Luttmer. (2009). "Who Misvotes? The Effect of Differential Cognition Costs on Election Outcomes", *American Economic Journal: Economic Policy*, 1(1): 229-257.

[35] Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. (1986). "Who Overreports Voting?" *American Political Science Review* 80:613–24.

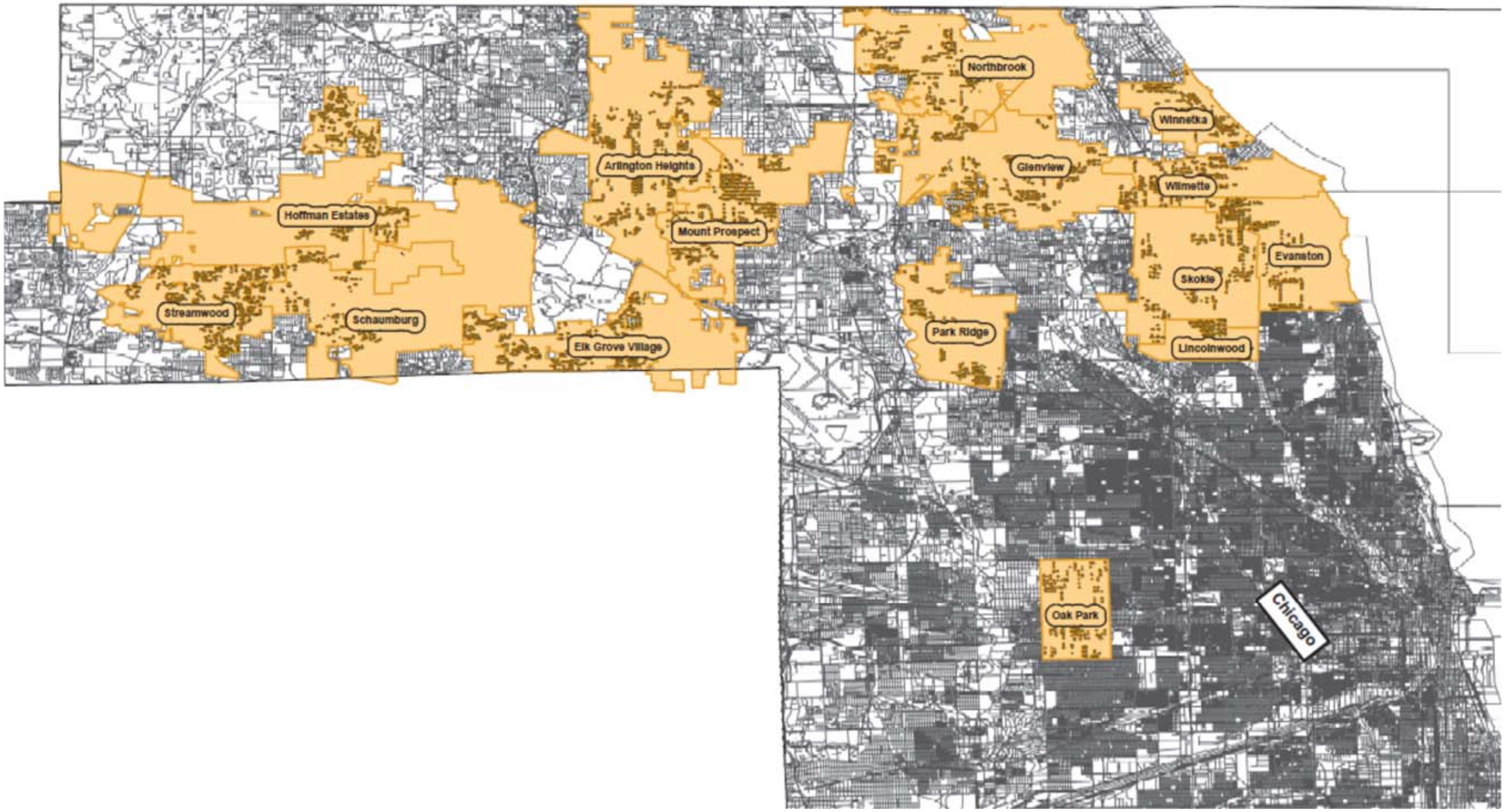
[36] Wooldridge, Jeffrey M. (2002). *Econometric analysis of cross section and panel data.* Cambridge, MA: The MIT Press.

Figure 1. Social-Image Value of Voting per Interaction (L = \$10)



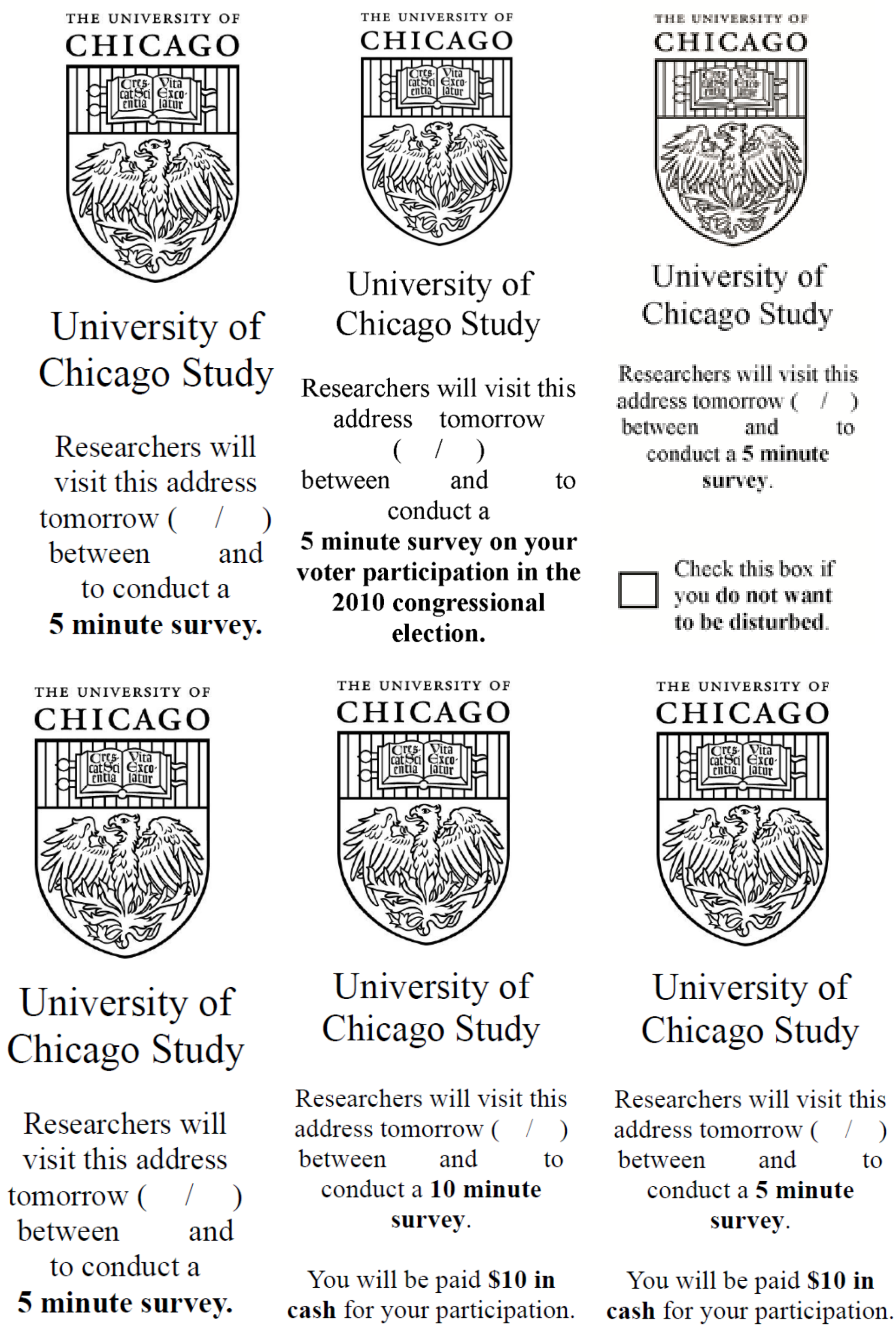
Note: Figure 1 plots the social-image value of voting due to the anticipation of being asked once, as a function of the net social image utility s_V-s_N and assuming a cost of lying L of \$10.

Figure 2. Area Surveyed



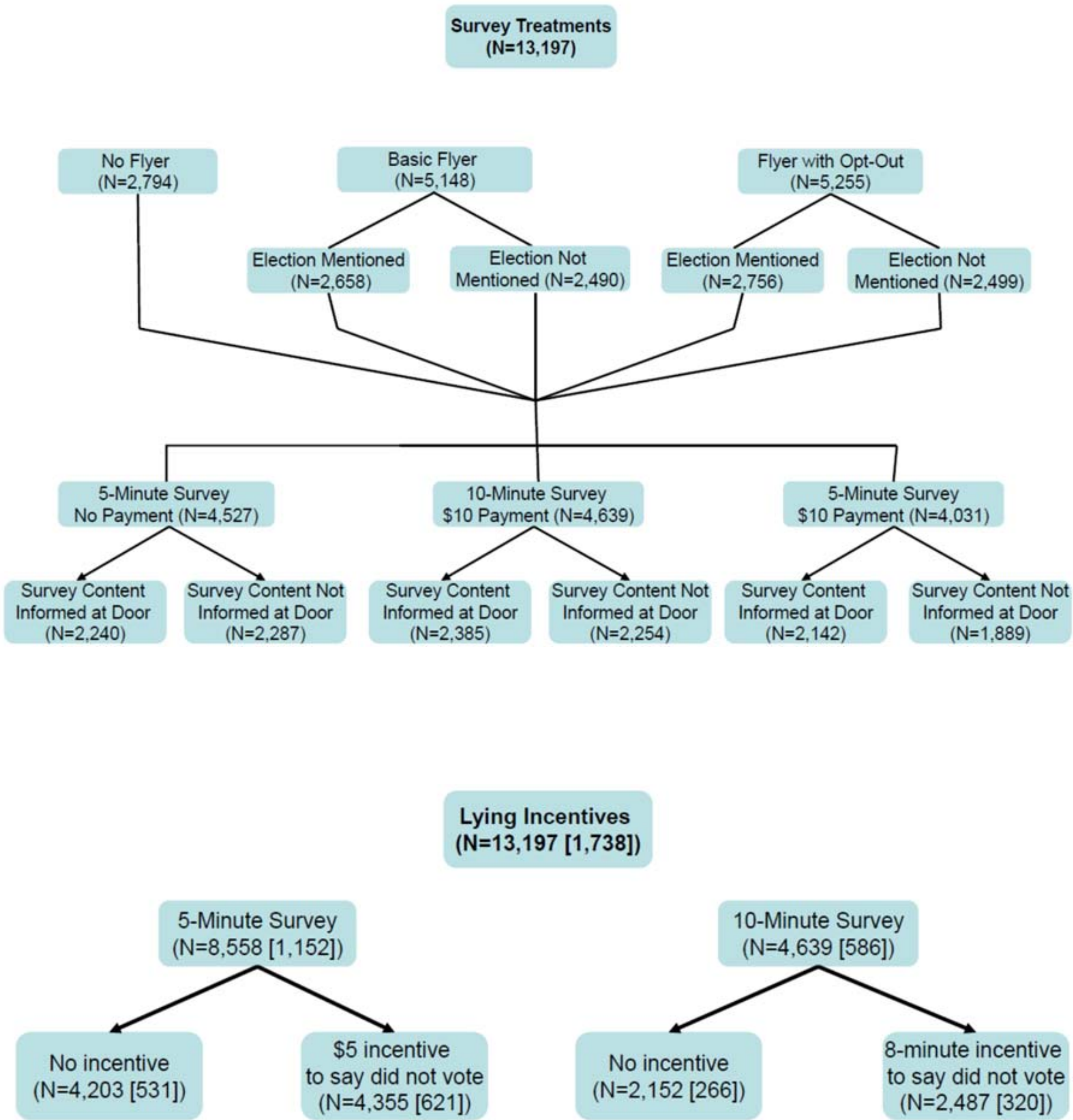
Note: Figure 2 displays in color (light grey in black/white) the towns visited in Cook County, Ill., as part of the door-to-door field experiment. The locations of the households visited within the towns are displayed in darker color. For reference, Chicago, which we did not visit, is located on the bottom-right corner.

Figure 3. Flyer Samples



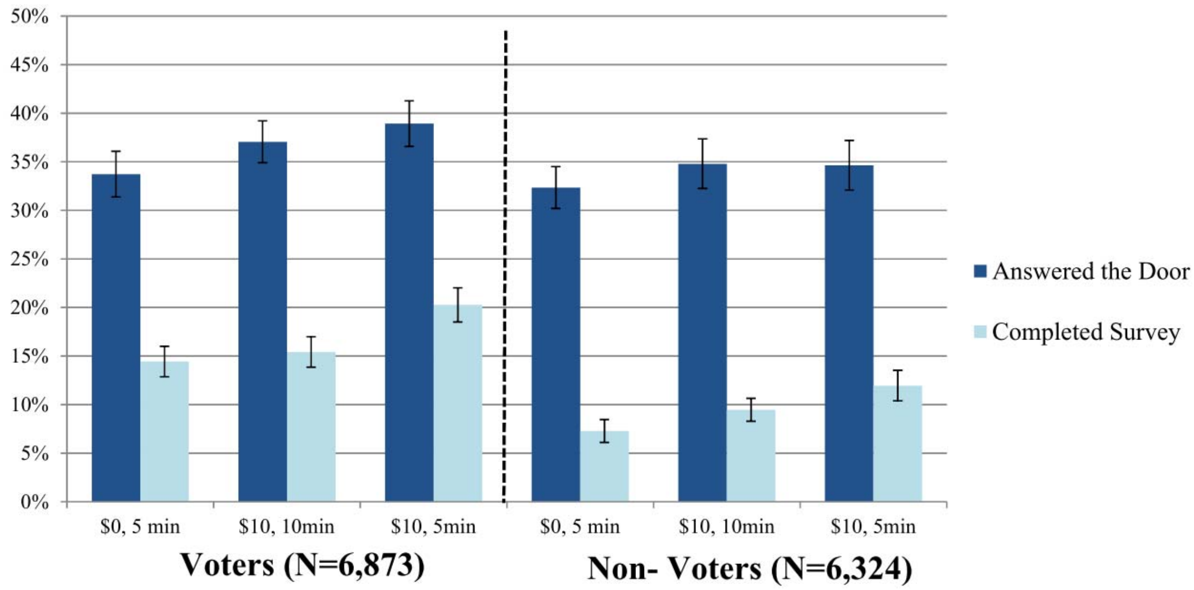
Note: The top three flyers are for (5-min., \$0) surveys in treatments Flyer (left), Flyer election (center), and Opt Out (right). The bottom three flyers are for Flyer treatments (5-min., \$0), (10-min., \$10), and (5-min., \$10).

Figure 4. Experimental Treatments



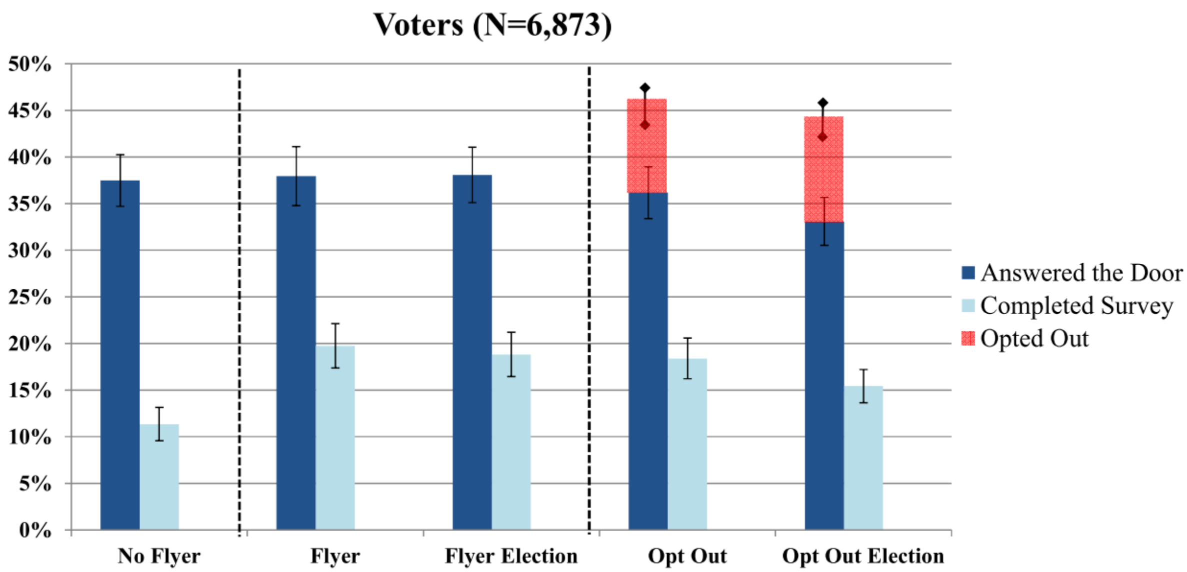
Note: Figure 4 presents the crossed experimental randomizations, with sample sizes in parentheses. On top are the five arms of the flyer treatment, crossed with whether respondents at the door are informed that the survey is about participation in the 2010 congressional election, crossed with survey duration and payment. At the bottom are the arms of the lying incentives, indicating both the initial sample size and [in square brackets] the sample size among individuals who responded to the survey. All arms are equally weighted and crossed.

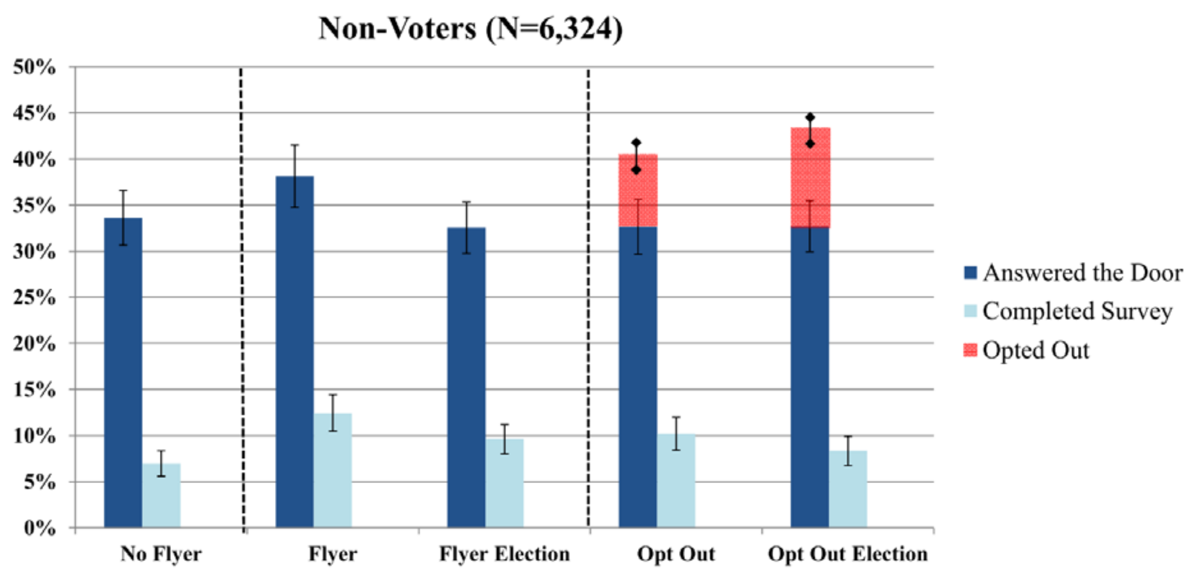
Figure 5. Response to Survey Duration and Payment



Note: Figure 5 presents the share of households answering the door and the (unconditional) share completing the survey across the three different combinations of payment and duration, separately for voting households and non-voting households. The averages are pooled across the different flyer treatments featured in Figure 4.

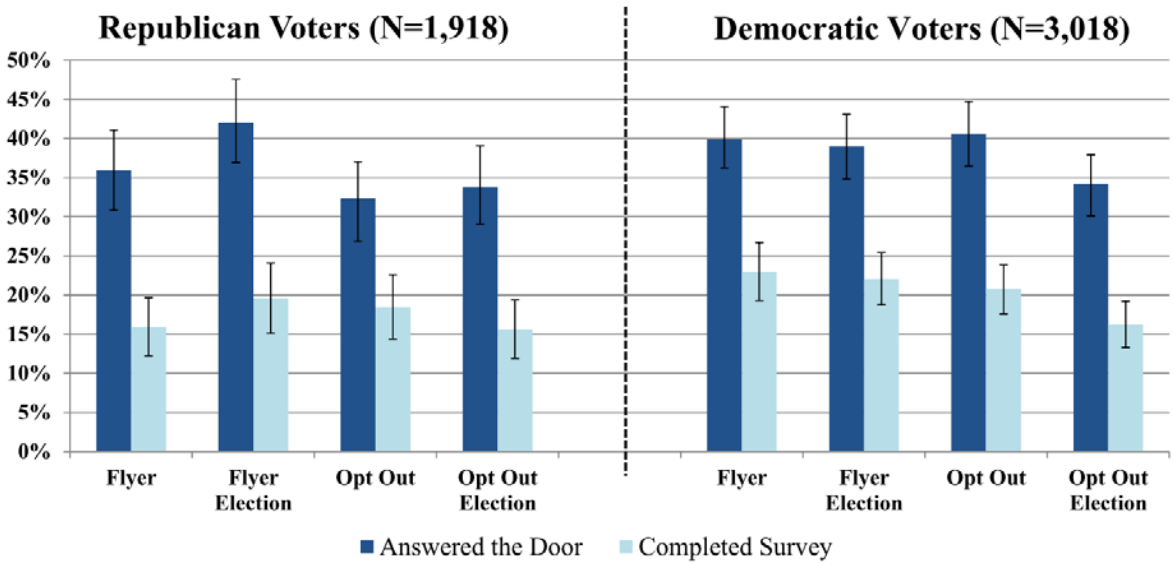
Figures 6. Response to Information about Election in Flyer





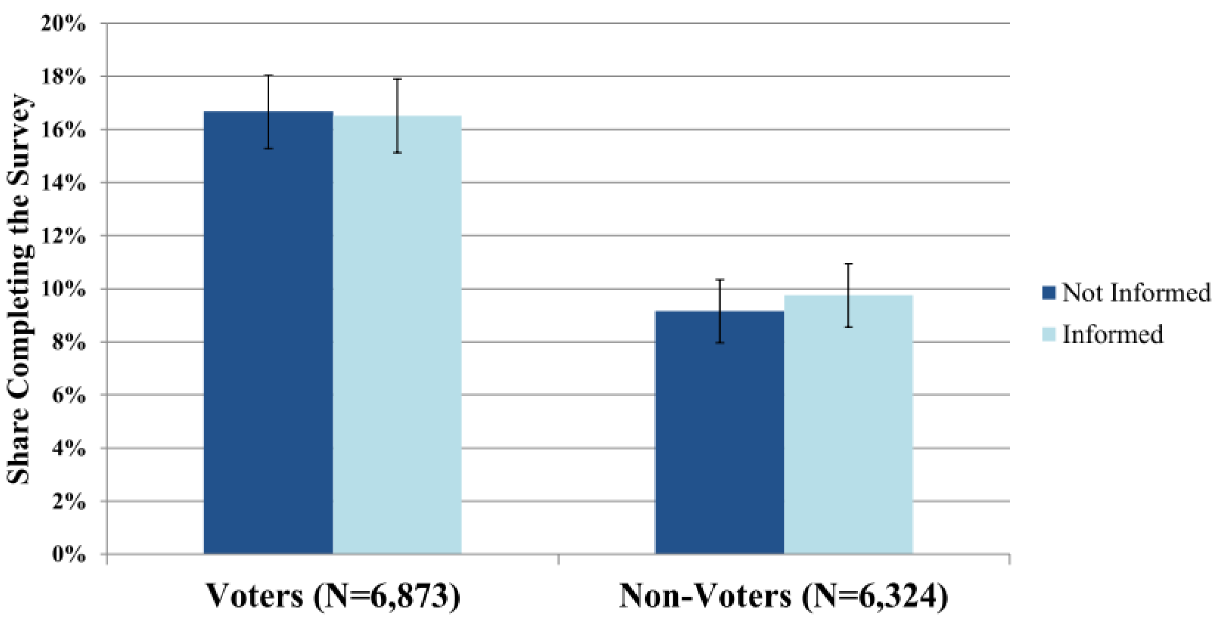
Note: Figure 6 presents the share of households answering the door, the (unconditional) share completing the survey, and (when applicable) the share opting out, separately for each of the five flyer treatments and separately for voting households and non-voting households. The averages are pooled across the three different payment and duration treatments featured in Figure 4.

Figure 7. Response to Flyer by Party Registration (*for voters*)



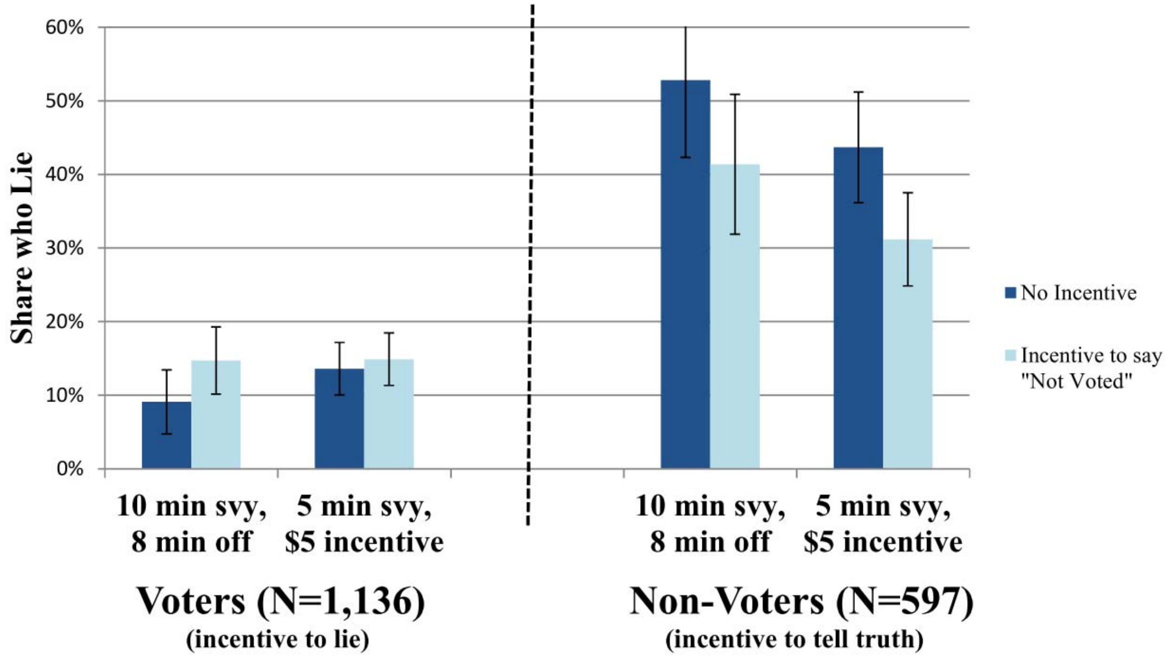
Note: Figure 7 presents the data from Figure 6a for *voting* households (omitting for space reasons the no-flyer treatment) split into two groups. In the left group, at least one household member voted at a Republican primary between 2004 and 2010. In the right group, at least one member voted at a Democratic primary between 2004 and 2010. Households with neither, or with voters participating in different party primaries, are not included.

Figure 8. Response to announcement of survey content at door



Note: Figure 8 presents the (unconditional) share of households completing the survey, separately for voting and non-voting households. The households in the *Not-Informed* treatment are not informed ex ante about the survey content at the door. The households in the *Informed* treatment are told at the door that the survey will be about their voter participation in the 2010 congressional election. The averages are pooled across the different flyer treatments featured in Figure 4.

Figure 9. Response to Lying Incentives



Note: Figure 9 presents the share of households completing the survey whose response to the question "Did you vote in the 2010 congressional election?" differs from the official voting record, denoted as "Share who Lie." The shares lying are compared across treatments with an incentive to say that one did not vote to the treatments with no such incentive. The incentives are designed to induce voters to lie and non-voters to tell the truth. The averages are pooled across the different flyer and payment treatments featured in Figure 4. The sample sizes refer to the subsamples who answered the survey including the voting question.

Figure 10. Social-Image Value of Voting as a function of Lying Cost

Figure 10a. Value of being asked once

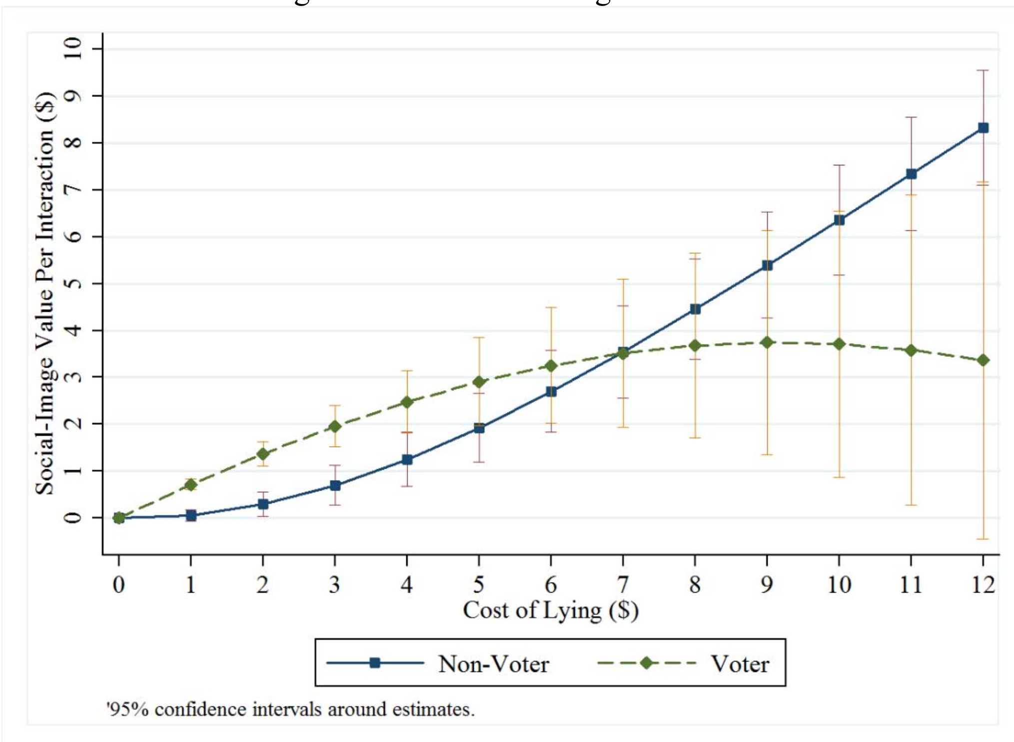
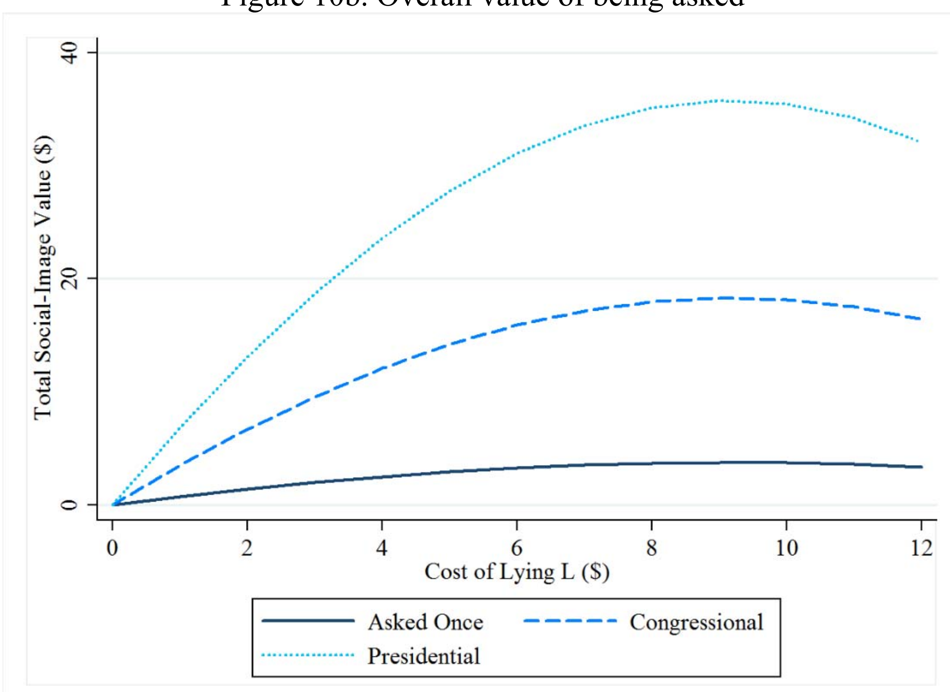


Figure 10b. Overall value of being asked



Note: Figure 10a plots the estimated value of voting due to being asked *once* for voters and non-voters at the benchmark estimate of the parameters, as a function of the lying cost *L* (which is not identified). The bars represent 95% confidence intervals. Figure 10b plots the same social-image value of voting for voters due to being asked once (same as Figure 10a), but also the total values for Presidential and Congressional elections. The latter estimates are the value due to being asked once, multiplied by the average number of times asked, taken from survey responses.

Figure 11. Flyer Samples for GOTV Treatment

THE UNIVERSITY OF CHICAGO

University of Chicago Study

Don't forget to vote in the 2012 Presidential Election.

Election Day is Tuesday, November 6, 2012.

THE UNIVERSITY OF CHICAGO

University of Chicago Study

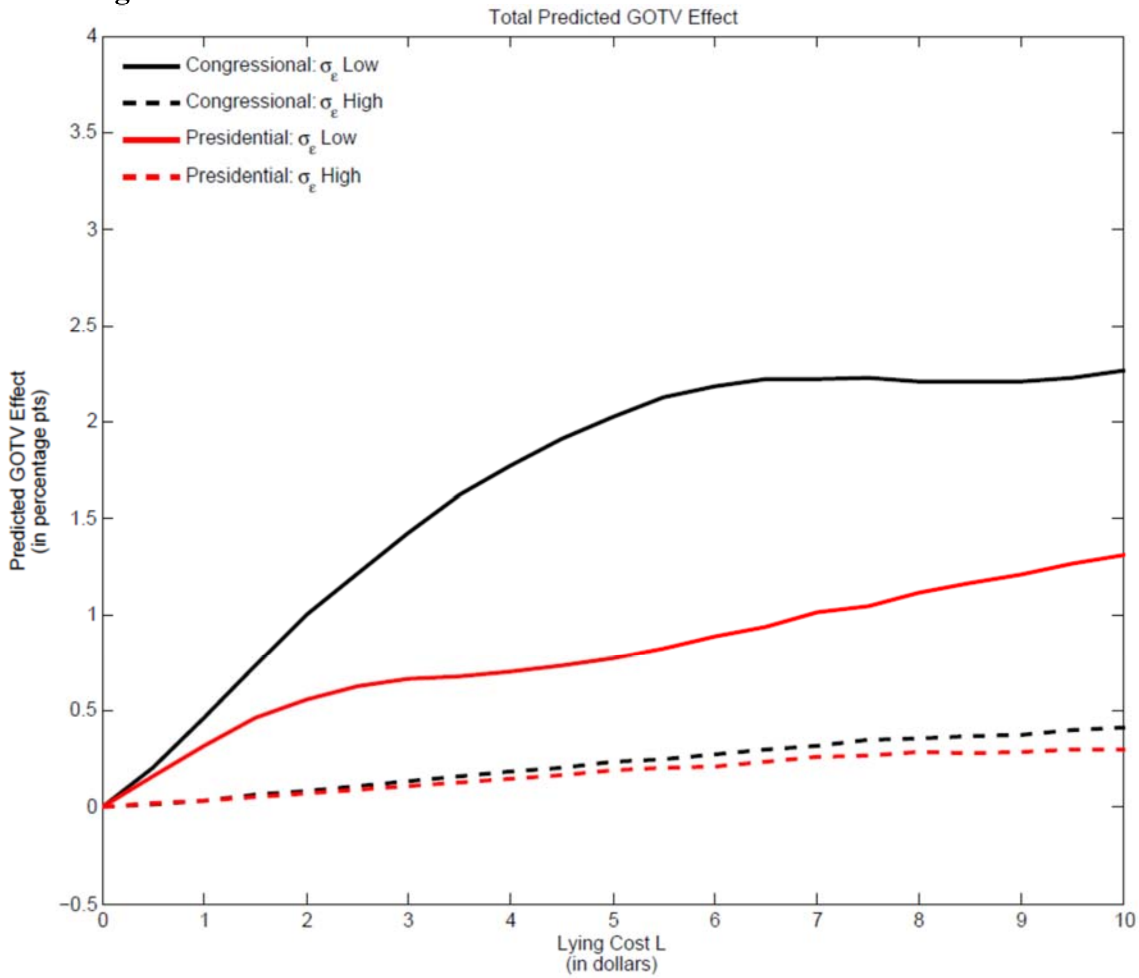
Researchers will contact you within three weeks of the election (between 11/7 and 11/27) to conduct a survey on your voter participation.

Don't forget to vote in the 2012 Presidential Election.

Election Day is Tuesday, November 6, 2012.

Note: Figure 11 shows the door-knob flyers used in the Get-Out-The-Vote treatments in the days before the 2012 presidential election. The left flyer is for the treatment with Voting Reminder, the right flyer is for the treatment with Announcement Will Ask About Voting. Flyers for the 2010 election are similarly styled.

Figure 12. Calibrated Get-Out-the-Vote Effect for Benchmark Estimates



Note: Figure 12 plots the implied get-out-the-vote turnout effect from announcing an additional visit to ask whether one voted, at the estimated benchmark parameters (Table 3). Since the estimated signaling value of voting depends on the assumed value of the lying cost L (which is not identified), we plot the estimated get-out-the-vote effect as a function of L. Further, we present two calibrations for the variance of the other reasons to vote, a low-variance scenario (the continuous line) and a high-variance scenario (the dotted line). Finally, in addition to the calibrations for the congressional elections (black line), we present calibrations for the presidential election (red line) assuming the same signaling value but a different number of times asked. Additional details are in Appendix D.

Table 1. Survey Treatments

Specification:	OLS Regressions							
Dependent Variable:	Indicator for Answering the Door				Indicator for Completing Survey			
Group:	Voters		Non-Voters		Voters		Non-Voters	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.3458***		0.3206***		0.0909***		0.0457***	
	(0.017)		(0.017)		(0.012)		(0.010)	
$10/10min Treatment	0.0337**	0.0364**	0.0251	0.0243	0.0109	0.0132	0.0226***	0.0231***
	(0.016)	(0.015)	(0.016)	(0.015)	(0.010)	(0.010)	(0.008)	(0.009)
$10/5min Treatment	0.0515***	0.0596***	0.0227	0.0204	0.0602***	0.0683***	0.0465***	0.0467***
	(0.017)	(0.017)	(0.015)	(0.015)	(0.013)	(0.013)	(0.009)	(0.009)
Simple Flyer Treatments	0.0145	0.0128	0.0306	0.0286	0.0907***	0.0960***	0.0522***	0.0496***
	(0.018)	(0.018)	(0.019)	(0.018)	(0.013)	(0.013)	(0.010)	(0.010)
Flyer Treatments with Opt-out	-0.0195	-0.0232	0.0055	0.0052	0.0673***	0.0695***	0.0354***	0.0325***
	(0.019)	(0.019)	(0.018)	(0.018)	(0.013)	(0.013)	(0.010)	(0.010)
Mention of Election in Flyer	-0.0158	-0.0143	-0.0276**	-0.0278**	-0.0200*	-0.0194*	-0.0236***	-0.0238***
	(0.013)	(0.013)	(0.014)	(0.014)	(0.011)	(0.011)	(0.008)	(0.008)
Voters Informed at Door of Election Topic					-0.0024	0.0001	0.0042	0.0047
					(0.009)	(0.009)	(0.008)	(0.008)
Omitted Treatment	**No Flyer, $0/5min Treatment**				**No Flyer, $0/5min, Not Informed Treatment**			
Fixed Effects for Solicitor, Date-Location, and Hour		X		X		X		X
R2	0.0032	0.0279	0.0018	0.0338	0.0116	0.0350	0.0080	0.0269
N	6,873	6,873	6,324	6,324	6,873	6,873	6,324	6,324

Notes: Estimates for a linear probability model with standard errors, clustered by solicitor-date, in parentheses. The table summarizes the result of three crossed treatments. The first is on duration and payment of the survey: $0/5min (the omited category), $10/10min, and $10/5min. The second is the flyer content: No flyer (the omitted category), Simple Flyer, Flyer with Opt-out. Each of the two flyer treatments is randomized into containing a mention of the election question in the flyer, or not. Hence, the coefficient on "Mention of Election in Flyer" captures the differential effect of the mention, compared to a simple flyer or a flyer with opt-out. Finally, we randomize whether, at the door, we announce the election question in the survey, with the omitted category being no mention. The regressions include fixed effects for the solicitor, for the date-town combination, and for the hour of day whenever indicated.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2. Lying Incentives

Specification:	OLS Regressions			
Dependent Variable:	Indicator for Lie (Stated Voting Does not Match Official Voting Record)			
Group:	Voters		Non-Voters	
	(1)	(2)	(3)	(4)
Panel A. All Survey Respondents				
Constant	0.1210***		0.4677***	
	(0.014)		(0.031)	
Time or Monetary Incentive To say Did Not Vote	0.0273	0.0225	-0.1204***	-0.1190***
	(0.020)	(0.019)	(0.040)	(0.040)
N	1,136	1,136	597	597
Panel B. $0, 5min. Treatments				
Constant	0.1479***		0.3971***	
	(0.028)		(0.061)	
5-Dollar Incentive to Say Did Not Vote	-0.0302	-0.0394	-0.0918	-0.1105
	(0.034)	(0.037)	(0.075)	(0.076)
N	329	329	163	163
Panel C. $10, 5min. Treatments				
Constant	0.1280***		0.4623***	0.3389**
	(0.024)		(0.046)	(0.152)
5-Dollar Incentive to Say Did Not Vote	0.0480	0.0550	-0.1452**	-0.1313**
	(0.036)	(0.034)	(0.059)	(0.065)
N	427	427	229	229
Panel D. $10, 10min. Treatments				
Constant	0.0909***		0.5281***	
	(0.022)		(0.053)	
8-Minute Incentive to Say Did Not Vote	0.0561*	0.0487	-0.1143	-0.0864
	(0.031)	(0.034)	(0.069)	(0.071)
N	380	380	205	205
Omitted Treatment	No incentive to say did not vote			
Fixed Effects for Location-Day		X		X

Notes: Estimates for a linear probability model with standard errors, clustered by solicitor-date, in parentheses. The regressions include fixed effects for location-day in Columns 2 and 4.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3. Minimum-Distance Estimates, Benchmark Results

	Benchmark Estimates	
Voting Parameters	Voter	Non-Voter
Mean Value of saying voted	-5.86	-4.61
(μ_V for voters, μ_V-L for non-voters)	(1.94)	(2.4)
Mean Value of saying didn't vote	-24.81	-4.23
(μ_N-L for voters, μ_N for non-voters)	(5.14)	(2.2)
Std. Dev. Of s_V and s_N	12.35	6.20
	(3.1)	(1.29)
Total Signaling Value of Voting, as a Function of Lying Cost (times asked: voters=5.13, non-voters=6.03)		
L=0	0.00	0.00
L=2	6.84	1.75
L=5	14.60	11.53
L=10	19.13	38.25
Average Utility From Being Asked About Voting	-4.63	-0.92
	(1.98)	(2.15)
Auxiliary Parameters	Voter	Non-Voter
Mean Utility (in $) of Doing 10-Minute Survey	-25.01	-31.62
	(3.92)	(5.51)
Std. Dev. of Utility of Doing Survey	29.90	28.09
	(7.57)	(6.73)
Value of Time of One-Hour Survey	65.04	19.09
	(14.86)	(8.58)
Social Pressure Cost (in $) of declining survey	1.76	0.06
	(1.2)	(0.45)
Elasticity of Home Presence (η)	0.13	2.86
	(0.09)	(20.2)
Probability of seeing the flyer	0.38	0.30
	(0.02)	(0.02)
Baseline Probability of being home	0.38	0.36
	(0.01)	(0.01)
SSE	135.15	

Notes: Estimates from minimum-distance estimator using the moments in Appendix Table 1 with weights given by the inverse of the diagonal of the variance-covariance matrix. The sample consists of 6,873 voting households and 6,324 non-voting households. A [non-]voting household is a household in which all registered voters did [not] vote in the 2010 congressional election. Standard errors are in parentheses. SSE reports the Weighted Sum of Squared Errors.

Table 4. Minimum-Distance Estimates, Identification with Different Sets of Moments

	Benchmark Model		Lying Rates Only		Sorting Only (Low Heterog.)		Sorting Only (High Heterog.)		No Survey Content Announcement at Door	
	(1)		(2)		(3)		(4)		(5)	
Voting Parameters	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter
Mean Value of saying voted	-5.86	-4.61	0.00	0.00	-5.25	-3.75	-8.7071	-19.791	-9.92	-13.60
(μ_V for voters, μ_V-L for non-voters)	(1.94)	(2.4)	(Assumed Value)		(3.02)	(1.74)	2.0275	2.4118	(3.46)	(4.48)
Mean Value of saying didn't vote	-24.81	-4.23	-22.68	0.72	-13.33	-3.17	-36.477	-17.887	-31.89	-12.41
(μ_N-L for voters, μ_N for non-voters)	(5.14)	(2.2)	(9.41)	(0.79)	(3.11)	(1.74)	2.9201	2.3937	(7.73)	(3.91)
Std. Dev. Of s_V and s_N	12.35	6.20	16.01	6.97	5.00	5.00	20.00	20.00	15.93	8.40
	(3.1)	(1.29)	(8.23)	(1.7)	(Assumed Value)		(Assumed Value)		(4.7)	(2.29)
Total Signaling Value of Voting, as a Function of Lying Cost										
L=0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L=2	6.84	1.75	6.55	1.24	6.19	1.90	6.54	0.03	6.40	0.65
L=5	14.60	11.53	14.40	9.73	7.99	13.17	14.79	2.61	13.98	7.37
L=10	19.13	38.25	21.16	34.91	-8.29	41.55	23.69	13.32	20.15	29.71
Average Utility From Being Asked About Voting	-4.63	-0.92	Unidentified		-4.81	-0.63	-6.27	-7.53	-7.96	3.46
	(1.98)	(2.15)			(3.03)	(1.72)	(2.07)	(2.25)	(8.24)	(3.71)
Auxiliary Parameters	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter
Mean Utility (in $) of Doing 10-Minute Survey	-25.01	-31.62	.	.	-37.53	-30.72	-29.23	-34.75	-21.29	-22.36
	(3.92)	(5.51)			(9.5)	(4.28)	(5.64)	(6.44)	(2.87)	(3.12)
Std. Dev. of Utility of Doing Survey	29.90	28.09	.	.	51.77	26.31	35.55	31.63	25.07	20.50
	(7.57)	(6.73)			(18.42)	(4.92)	(10.26)	(8.01)	(5.4)	(3.64)
Value of Time of One-Hour Survey	65.04	19.09	14.63		80.44	30.15	94.68	35.17	78.36	27.01
	(14.86)	(8.58)	(10.41)		(18.15)	(15.33)	(23.37)	(25.29)	(16.09)	(11.28)
SSE	135.15		5.32		121.64		154.78		103.25	

Notes: Estimates from minimum-distance estimator using the moments in Appendix Table 1 with weights given by inverse of diagonal of variance-covariance matrix. The sample consists of 6,873 voting households and 6,324 non-voting households. Standard errors are in parentheses. SSE reports the Weighted Sum of Squared Errors. The **Benchmark Model** uses all the moments - answering the door, opting out, completing the survey, and lying rates under different incentives. The **Lying Rates Only** model uses *only* the lying rate moments under different incentives to lie, and drops the moments for answering the door, opting out and completing the survey. The **Sorting Only** models use as moments the shares of voters and non-voters answering the door, opting out and doing the survey under the various survey length, payment and advance warning / opt out treatments, the survey content announcement at door treatments, *as well as the baseline lying rate* - i.e. the model does not exploit the experimental variation in incentives to say one did not vote. Within the Sorting Only model, we vary the assumed heterogeneity in signaling utility, which is unidentified without the lying incentives. The **No Survey Content Announcement at Door** model drops the moments for the treatment varying the announcement at the door.

Table 5. Minimum-Distance Estimates, Robustness Table

	Benchmark Model: Sorting and Lying		Allowing Utility from Talking About Politics		Assuming 10% Non-Registered in Voting Households		Assuming 10% Misclassified	
	(1)		(2)		(3)		(4)	
Voting Parameters	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter
Mean Value of saying voted	-5.86	-4.61	0.00		-5.18	-4.52	-5.13	-4.08
(μ_V for voters, μ_V-L for non-voters)	(1.94)	(2.4)	(Assumed)		(1.95)	(2.43)	(1.92)	(2.77)
Mean Value of saying didn't vote	-24.81	-4.23	-18.99	0.39	-27.46	-4.06	-26.79	-2.59
(μ_N-L for voters, μ_N for non-voters)	(5.14)	(2.2)	(3.79)	(0.64)	(6.19)	(2.18)	(5.73)	(2.47)
Std. Dev. Of s_V and s_N	12.35	6.20	12.38	6.21	11.40	6.27	10.52	5.63
	(3.1)	(1.29)	(3.06)	(1.29)	(3.29)	(1.27)	(2.89)	(1.25)
Utility of Talking about Politics			-5.87	-5.39				
			(1.94)	(2.4)				
Total Signaling Value of Voting, as a Function of Lying Cost (times asked: voters=5.13, non-voters=6.03)								
L=0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L=2	6.84	1.75	6.84	1.75	8.11	1.65	8.34	0.61
L=5	14.60	11.53	14.61	11.52	18.17	11.24	18.69	9.67
L=10	19.13	38.25	19.17	38.24	26.97	37.80	27.40	36.48
Average Utility From Being Asked	-4.63	-0.92	-4.61	-0.91	-4.57	-0.75	-7.48	-0.16
About Voting	(1.98)	(2.15)	(1.98)	(2.14)	(2.)	(2.12)	(4.13)	(2.3)
Survey Parameters	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter
Value of Time of One-Hour Survey	65.04	19.09	65.14	19.13	76.13	17.11	75.91	13.39
	(14.86)	(8.58)	(14.68)	(8.58)	(16.75)	(8.63)	(16.14)	(9.46)
SSE	135.15		135.15		132.31		132.63	

Notes: Estimates from minimum-distance estimator using the moments in Appendix Table 1 with weights given by the inverse of the diagonal of the variance-covariance matrix. The sample consists of 6,873 voting households and 6,324 non-voting households. Standard errors are in parentheses. SSE reports the Weighted Sum of Squared Errors. The **Allowing Utility from Talking About Politics** model adds a parameter capturing taste for talking about politics, independent of whether the respondent states that she voted or not. This requires either the mean value of saying one voted or of of saying one did not vote to be fixed. We normalize the mean utility of saying one voted to zero. The **Assuming 10% Non-Registered in Voting Households** model assumes that 10% of households classified as all-voters from registration data actually include non-registered adults who are non-voters by definition, but are misclassified as voters by us. The **Assuming 10% Misclassified** model assumes measurement error of 10% in the voter turnout data or our matching of it to households in our sample. Thus, 10% of voting households are assumed to be non-voters, and vice versa.

Table 6. Full Estimation - Subset with Intermediate propensity to vote

	Assuming Voters = Non-Voters		Assuming Voters = Non-Voters Partially		Voters and Non-Voters Different	
	(1)		(2)		(3)	
Voting Parameters	Voters = Non-Voters		Voters = Non-Voters		Voters	Non-Voter
Mean Utility from *truthfully* saying voted (μ_V)	-1.06		-2.34		-4.85	1.06
	(3.13)		(2.75)		(4.14)	(4.76)
Mean Utility from *truthfully* saying didn't vote (μ_N)	-5.56		-6.79		-12.45	-4.12
	(3.21)		(2.9)		(6.93)	(4.64)
Cost of Lying (L)	4.63		5.17		5.00	5.00
	(1.08)		(1.29)		(Assumed)	(Assumed)
Std. Dev. Of Sv and Sn	7.34		8.68		12.68	5.93
	(1.62)		(2.05)		(4.86)	(1.47)
	Voters	Non-Voters	Voters	Non-Voters	Voters	Non-Voters
Total Signaling Value of Voting (times asked: voters=5.50, non-voters=5.85)	8.28	8.80	7.82	8.32	8.92	12.88
Average Utility From Being Asked About Voting	0.02	-1.48	-0.82	-2.25	-2.31	-0.69
Survey Parameters	Voters = Non-Voters		Voters	Non-Voter	Voters	Non-Voter
Mean Utility (in $) of Doing 10-Minute Survey	-46.62		-34.81	-53.98	-37.78	-47.69
	(12.99)		(10.14)	(21.13)	(12.08)	(21.85)
Std. Dev. of Utility of Doing Survey	51.63		34.73	61.14	38.30	52.08
	(17.49)		(14.42)	(29.85)	(16.88)	(29.13)
Value of Time of One-Hour Survey	28.85		55.09	12.58	70.92	15.36
	(9.99)		(18.72)	(17.12)	(28.42)	(11.53)
SSE	132.49		100.79		97.20	

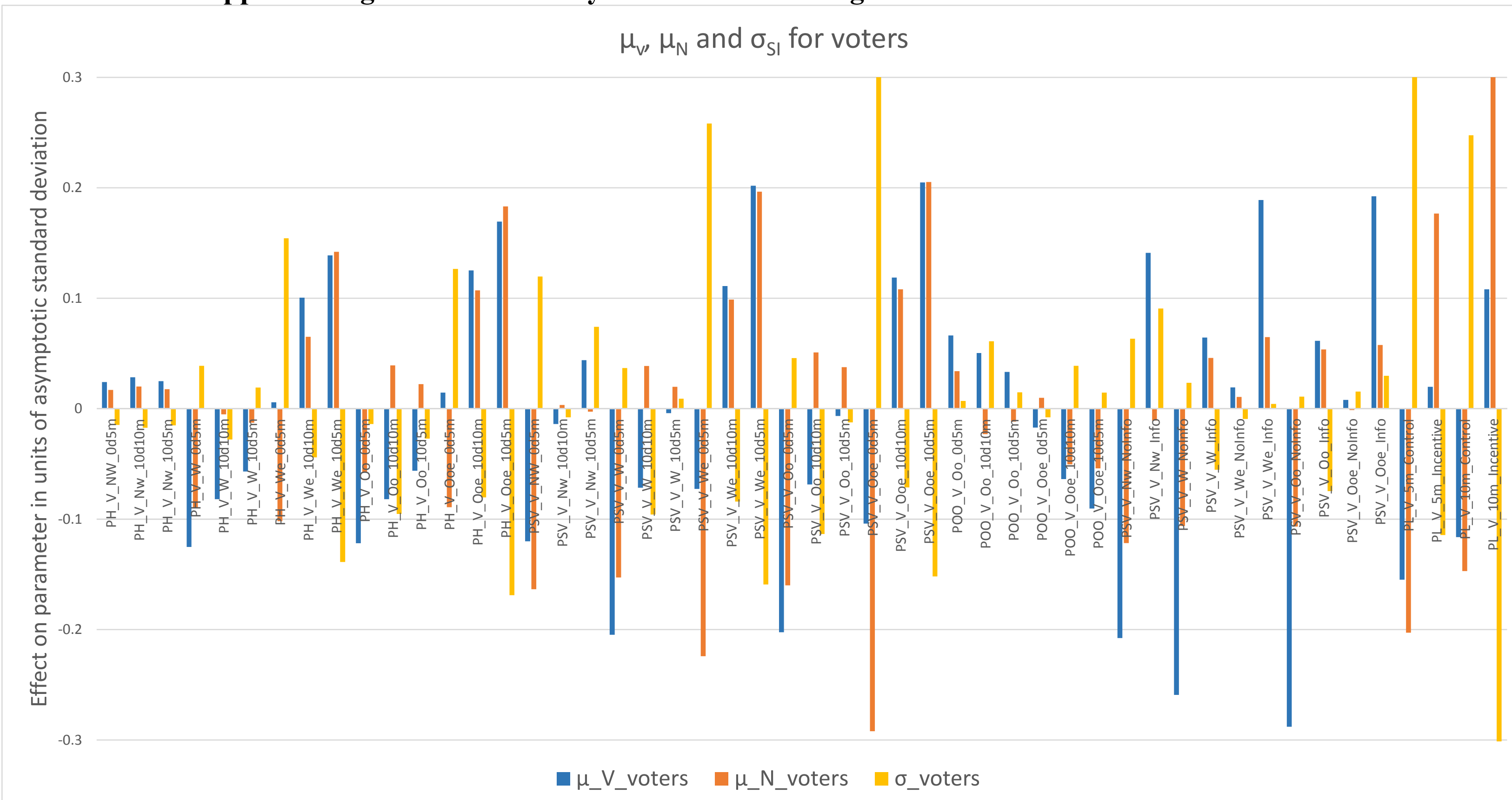
Notes: Estimates from minimum-distance estimator with weights given by the inverse of the diagonal of the variance-covariance matrix. Standard errors are in parentheses. SSE reports the Weighted Sum of Squared Errors. The sample includes only households with intermediate voting propensity, excluding always-voters and never-voters. Specifically, for each registerd voter in the household we compute the predicted probability of voting in the Nov. 2010 congressional elections using a probit model with indicators for participation in the following elections: March 2004 (primary), Nov. 2004, Feb. 2005, March 2006 (primary), Nov. 2006, April 2007, Feb. 2008 (primary), Nov. 2008, April 2009, and Feb. 2010 (primary). We then keep only the 5,901 households in which all registered voters have a predicted probability of voting between 25 percent and 75 percent. We re-estimate the moments on this sample. Column 1 reports estimates assuming that voters and non-voters in this sample have the same parameters. Column 2 reports estimates assuming that voters and non-voters in this sample share the same voting parameters, but may differ in their other parameters. The restrictions imposed in Columns 1 and 2 allow us to identify also the cost of lying. Finally, in Column 3 we allow voters and non-voter to differ with respect to all parameters, as in the benchmark specification but on a restricted sample.

Table 7. Results for Get-Out-The-Vote Treatments

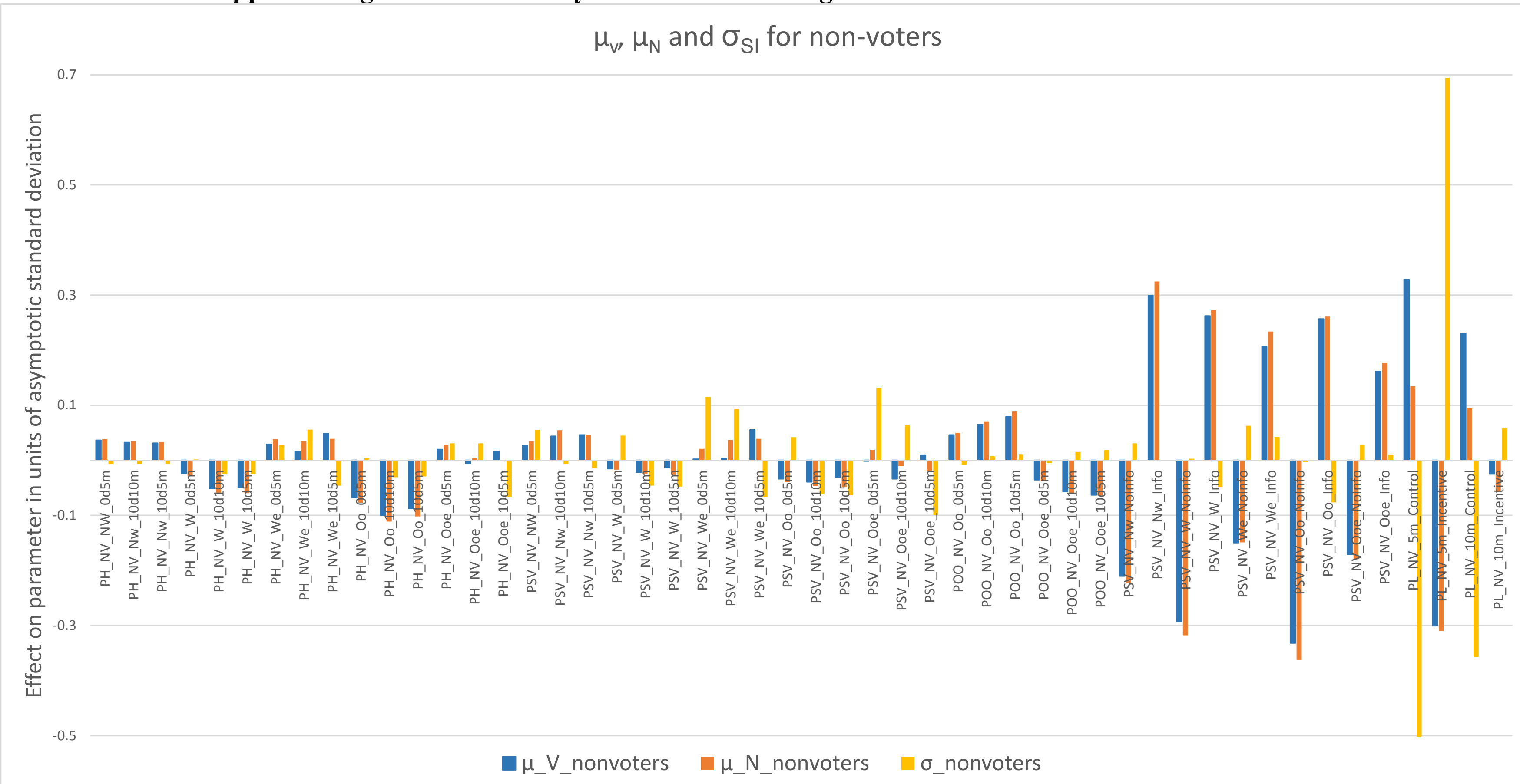
Specification:	OLS Regressions			
Dependent Variable:	Indicator for Voting			
Election:	Congressional Elections in Nov. 2010		Presidential Elections in Nov. 2012	
	(1)	(2)	(3)	(4)
Constant	0.6000***		0.7312***	
	(0.0109)		(0.0033)	
Flyer with Voting Reminder	-0.0020	-0.0031	0.0060	0.0046
	(0.0152)	(0.0083)	(0.0056)	(0.0034)
Flyer with Announcement Will Ask About Voting	0.0120	0.0102	0.0023	0.0056
	(0.0157)	(0.0084)	(0.0056)	(0.0034)
Omitted Treatment	No Flyer		No Flyer	
Control for past Voting since 2004		X		X
Difference (Flyer Will Ask - Flyer Reminder)	0.0140	0.0133	-0.0037	0.0010
p-value for test of equality, 2-sided	p=0.365	p=0.120	p=0.561	p=0.811
p-value for test of equality, 1-sided	p=0.182	p=0.060*		p=0.405
R2	0.0001	0.4024	0.0000	0.3251
N	N = 31,306	N = 31,304	N = 93,805	N = 93,805

Notes: Estimates for a linear probability model with standard errors, clustered by flyering route, in parentheses. The omitted treatment is the No-Flyer condition. The regressions in Columns 2 and 4 include indicators for participation in the following elections: March 2004 (primary), Nov. 2004, Feb. 2005, March 2006 (primary), Nov. 2006, April 2007, Feb. 2008 (primary), Nov. 2008, April 2009, and Feb. 2010 (primary). In addition, Column 4 includes indicators for participation in the Nov. 2010, April 2011, and March 2012 (primary) elections.
* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix Figure 1a. Sensitivity of Estimated Voting Parameters to the Moments: Voters



Appendix Figure 1b. Sensitivity of Estimated Voting Parameters to the Moments: Non-Voters



Note: Appendix Figures 1a and 1b present the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). The plotted lines indicate the local sensitivity of the given parameter estimate to an individual moment. A positive bar indicates that a (local) increase in the moment would increase the estimated value of the parameter. Higher bars indicate more influential moments for the identification of the parameter. For each moment, we plot the influence estimate for the mean value of saying one voted (in blue), the mean value of saying one did not vote (in orange) and the standard deviation of such values (in orange) respectively for voters (Appendix figure 1a) and for non-voters (Appendix Figure 1b).

Appendix Table 1. Empirical Moments and Predicted Moments at Benchmark Estimates

Voters

	Empirical Mom.	Predicted Mom.		Empirical Mom.	Predicted Mom.
Flyers: Home Presence	**P(Home)**		**Flyers: Survey Compl.**	**P(Do Survey)**	
No Flyer, $0,5min	0.3499	0.3799	No Flyer, $0,5min	0.1145	0.1023
No Flyer, $10,10min	0.4081	0.3799	No Flyer, $10,10min	0.0993	0.1215
No Flyer, $10,5min	0.3827	0.3799	No Flyer, $10,5min	0.127	0.1398
Survey Flyer, $0,5min	0.3491	0.3740	Survey Flyer, $0,5min	0.1624	0.1499
Survey Flyer, $10,10min	0.3890	0.3893	Survey Flyer, $10,10min	0.1757	0.1797
Survey Flyer, $10,5min	0.4066	0.4041	Survey Flyer, $10,5min	0.2537	0.2083
Election Flyer, $0,5min	0.3686	0.3583	Election Flyer, $0,5min	0.1665	0.1383
Election Flyer, $10,10min	0.3701	0.3807	Election Flyer, $10,10min	0.1691	0.1757
Election Flyer, $10,5min	0.4233	0.4030	Election Flyer, $10,5min	0.2332	0.2123
Survey Opt Out, $0,5min	0.3586	0.3263	Survey Opt Out, $0,5min	0.1557	0.1497
Survey Opt Out, $10,10min	0.3553	0.3447	Survey Opt Out, $10,10min	0.172	0.1796
Survey Opt Out, $10,5min	0.3794	0.3625	Survey Opt Out, $10,5min	0.2213	0.2083
Election Opt Out, $0,5min	0.2643	0.3065	Election Opt Out, $0,5min	0.1099	0.1337
Election Opt Out, $10,10min	0.3457	0.3318	Election Opt Out, $10,10min	0.1506	0.1710
Election Opt Out, $10,5min	0.3926	0.3580	Election Opt Out, $10,5min	0.2034	0.2086
Opting Out	**P(Opt Out)**		**Informed at Door of Survey Content**	**P(Do Survey)**	
Opt Out, $0,5min	0.1330	0.1061	Not Informed, No Flyer	0.1097	0.1289
Opt Out, $10,10min	0.0874	0.0988	Informed, No Flyer	0.115	0.1135
Opt Out, $10,5min	0.0681	0.0918	Not Informed, Survey Flyer	0.1947	0.1912
Election Opt Out, $0,5min	0.1358	0.1108	Informed, Survey Flyer	0.1993	0.1673
Election Opt Out, $10,10min	0.1043	0.1042	Not Informed, Election Flyer	0.1844	0.1802
Election Opt Out, $10,5min	0.0864	0.0979	Informed, Election Flyer	0.1898	0.1707
			Not Informed, Opt Out	0.1814	0.1883
Lying	**P(Lie | Survey)**		Informed, Opt Out	0.1764	0.1702
5 min survey, control	0.1360	0.1042	Not Informed, Election Opt Out	0.1573	0.1759
5 min survey, $5 incent.	0.1489	0.1440	Informed, Election Opt Out	0.1507	0.1663
10 min survey, control	0.0909	0.1044			
10 min survey, 8 min incent.	0.1471	0.1464			

Non-Voters

	Empirical Mom.	Predicted Mom.		Empirical Mom.	Predicted Mom.
Flyers: Home Presence	**P(Home)**		**Flyers: Survey Compl.**	**P(Do Survey)**	
No Flyer, $0,5min	0.3674	0.3581	No Flyer, $0,5min	0.0827	0.0503
No Flyer, $10,10min	0.3298	0.3581	No Flyer, $10,10min	0.048	0.0781
No Flyer, $10,5min	0.3046	0.3581	No Flyer, $10,5min	0.0804	0.0846
Survey Flyer, $0,5min	0.3664	0.3376	Survey Flyer, $0,5min	0.0894	0.0751
Survey Flyer, $10,10min	0.3524	0.3569	Survey Flyer, $10,10min	0.1163	0.1172
Survey Flyer, $10,5min	0.413	0.3614	Survey Flyer, $10,5min	0.1664	0.1271
Election Flyer, $0,5min	0.3049	0.3238	Election Flyer, $0,5min	0.0718	0.0633
Election Flyer, $10,10min	0.3333	0.3387	Election Flyer, $10,10min	0.1035	0.1016
Election Flyer, $10,5min	0.3247	0.3487	Election Flyer, $10,5min	0.1168	0.1172
Survey Opt Out, $0,5min	0.2869	0.2935	Survey Opt Out, $0,5min	0.0642	0.0747
Survey Opt Out, $10,10min	0.3565	0.3169	Survey Opt Out, $10,10min	0.1148	0.1166
Survey Opt Out, $10,5min	0.3184	0.3224	Survey Opt Out, $10,5min	0.1226	0.1265
Election Opt Out, $0,5min	0.2785	0.2796	Election Opt Out, $0,5min	0.0429	0.0633
Election Opt Out, $10,10min	0.3511	0.2986	Election Opt Out, $10,10min	0.093	0.1018
Election Opt Out, $10,5min	0.346	0.3096	Election Opt Out, $10,5min	0.1211	0.1174
Opting Out	**P(Opt Out)**		**Informed at Door of Survey Content**	**P(Do Survey)**	
Opt Out, $0,5min	0.1058	0.0914	Not Informed, No Flyer	0.0736	0.0720
Opt Out, $10,10min	0.063	0.0830	Informed, No Flyer	0.0647	0.0700
Opt Out, $10,5min	0.0559	0.0810	Not Informed, Survey Flyer	0.1245	0.1099
Election Opt Out, $0,5min	0.1456	0.0918	Informed, Survey Flyer	0.1223	0.1029
Election Opt Out, $10,10min	0.0966	0.0837	Not Informed, Election Flyer	0.0855	0.0947
Election Opt Out, $10,5min	0.079	0.0818	Informed, Election Flyer	0.1061	0.0933
			Not Informed, Opt Out	0.0858	0.1099
Lying	**P(Lie | Survey)**		Informed, Opt Out	0.1111	0.1020
5 min survey, control	0.4335	0.4815	Not Informed, Election Opt Out	0.0834	0.0949
5 min survey, $5 incent.	0.3088	0.3055	Informed, Election Opt Out	0.084	0.0935
10 min survey, control	0.5281	0.4817			
10 min survey, 8 min incent.	0.4138	0.3781			

Notes: The Table presents the empirical moments and the predicted moments from a minimum-distance estimator. The empirical moments are obtained as regression estimates after controlling for the randomization fixed effects. The minimum-distance estimates are in Table 3, shown are the predicted moments at those parameter values.

Appendix Table 2. Minimum-Distance Estimates, Robustness Table - Appendix

	Benchmark Model		Identity Weighting Matrix		Pos. correlated s_V and s_N (ρ = 0.5)		Neg. correlated s_V and s_N (ρ = -0.5)		Large Sorting Elasticity (η = 1)		Small Sorting Elasticity (η = 0.02)	
	(1)		(2)		(3)		(4)		(5)		(6)	
Voting Parameters	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter
Mean Value of saying voted	-5.86	-4.61	-5.83	-5.98	-6.30	-5.20	-5.69	-4.37	-6.00	-4.61	-15.18	-6.18
(μ_V for voters, μ_V-L for non-voters)	(1.94)	(2.4)	(2.05)	(2.68)	(2.15)	(2.49)	(1.87)	(2.28)	(1.89)	(2.3)	(10.47)	(3.12)
Mean Value of saying didn't vote	-24.81	-4.23	-24.23	-5.63	-22.49	-5.03	-26.58	-3.81	-24.70	-4.21	-34.23	-5.43
(μ_N-L for voters, μ_N for non-voters)	(5.14)	(2.2)	(5.23)	(2.38)	(4.71)	(2.3)	(5.41)	(2.)	(4.94)	(2.04)	(12.47)	(2.78)
Std. Dev. Of s_V and s_N	12.35	6.20	11.72	6.34	13.87	7.95	11.66	5.44	12.44	6.25	10.87	7.30
	(3.1)	(1.29)	(3.09)	(1.43)	(3.47)	(1.48)	(2.88)	(1.17)	(3.01)	(1.27)	(3.54)	(1.8)
Total Signaling Value of Voting, as a Function of Lying Cost (times asked: voters=5.13, non-voters=6.03)												
L=0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L=2	6.84	1.75	6.94	1.74	7.10	2.18	6.67	1.45	6.74	1.72	7.50	1.15
L=5	14.60	11.53	14.73	11.38	14.63	13.01	14.47	10.48	14.31	11.42	16.18	9.28
L=10	19.13	38.25	18.84	37.91	16.36	40.76	20.30	36.32	18.53	38.06	21.42	33.89
Average Utility From Being Asked	-4.63	-0.92	-4.72	-2.22	-5.47	-1.94	-4.12	-0.33	-4.70	-0.88	-14.38	-1.68
About Voting	(1.98)	(2.15)	(2.1)	(2.3)	(2.19)	(2.25)	(1.92)	(1.93)	(1.94)	(1.98)	(10.5)	(2.69)
Survey Parameters	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter	Voter	Non-Voter
Value of Time of One-Hour Survey	65.04	19.09	61.72	14.54	50.57	18.36	71.97	19.61	67.95	19.30	35.12	17.11
	(14.86)	(8.58)	(14.84)	(9.6)	(12.38)	(8.03)	(15.59)	(8.87)	(14.45)	(8.61)	(12.08)	(10.44)
Elasticity of Home Presence (η)	0.13	2.86	0.14	9.31	0.11	1.45	0.14	8.03	1		0.020	
	(0.09)	(20.2)	(0.11)	(637.01)	(0.08)	(11.35)	(0.1)	(379.73)	(Assumed)		(Assumed)	
SSE	135.15		3.82		145.44		130.15		139.709		172.17	

Notes: Estimates from minimum-distance estimator using the moments in Appendix Table 1 with weights given by inverse of diagonal of variance-covariance matrix (except for Column 2). The sample size for the field experiment consists of 6,873 voting households and 6,324 non-voting households. The Identity Weighting Matrix model uses equally weighted moments. The estimation assumes bivariate normal distribution for signaling values of saying voted/didn't vote, with correlation zero in models 1, 5, and 6, correlation +0.5 in model 3, and -0.5 in model 4. Models 5 and 6 exogenously fix the sorting elasticity to high and low values respectively, and force the same elasticity for voters and non-voters. Standard errors are in parentheses. SSE reports the Weighted Sum of Squared Errors. All models use the entire set of empirical moments used in the benchmark estimation.

A Appendix A - Mathematical Appendix – For Online Publication

Proof of Propositions 1 and 2. We consider first the probability of being at home. As discussed in the text, the probability of being at home will be: (i) h_0 in the absence of flyer, or if the person does not see the flyer; (ii) $h^{i*} = \max\left[\min\left[h_0 + \eta^i \max\left(s^i + m - c^i, -S^i\right), 1\right], 0\right]$ if the person saw a survey flyer, and (iii) $h^{i*} = \max\left[\min\left[h_0 + \eta^i \max\left(s^i + m - c^i + z^i, -S^i\right), 1\right], 0\right]$ if the person saw an election flyer. Under *Pride in Voting,* $z^v = \max\left(s_V^v, s_N^v - L^v\right) \geq s_V^v$ is positive. Hence, h^* will be at least as high under FE than under F for voters. Conversely, under *Stigma from Not Voting,* $z^{nv} = \max\left(s_V^{nv} - L^{nv}, s_N^{nv}\right)$ is negative, and hence h^* will be lower under FE than under F for non-voters. Under opt-out, a person who sees the flyer will opt out (and hence set $h^* = 0$) if $s^i + m - c^i < 0$ under OO and if $s^i + m - c^i + z^i < 0$ under OOE. Under *Pride in Voting,* z^v is positive; hence, for any set of parameters, if the person opts out under OOE, she will also do so under OO (but not the converse). Hence, for any given set of parameters treatment, the probability of being at home is lower under OO than under OOE and thus $P\left(H\right)_{OOE}^v \geq P\left(H\right)_{OO}^v$. Conversely, under *Stigma from Not Voting,* z^{nv} is negative so the converse result applies and $P\left(H\right)_{OOE}^{nv} \leq P\left(H\right)_{OO}^{nv}$ follows.

Turning to the probability of answering a survey, conditional on answering the door, an individual will agree to the survey if $s^i + m - c^i + z^i \geq -S^i$ assuming she knows that the survey has an election topic and if $s^i + m - c^i \geq -S^i$ in case she does not know. By the same token as above, holding constant the selection into being at home, the person will be more likely to complete the survey if informed about the election topic under *Pride* and if not informed under *Stigma*. Hence, the conclusion $P\left(SV\right)_I^v \geq P\left(SV\right)_{NI}^v$ under *Pride* and $P\left(SV\right)_I^{nv} \leq P\left(SV\right)_{NI}^{nv}$ under *Stigma* hold (remember that the treatments I and NI take place after the sorting decision).

To consider the effect of F and FE on $P\left(SV\right)$ we need to take into account the selection into answering the door. We consider separately the following four exhaustive cases: (i) $max(s^i + m - c^i + z^i, s^i + m - c^i) < -S^i$. In this case, $P\left(SV\right) = 0$ under any condition; (ii) $\min(s^i + m - c^i + z^i, s^i + m - c^i) \geq -S^i$. In this case, the person will complete the survey conditional on being at home, so $P\left(H\right) = P\left(SV\right)$, and the comparison follows from the results above on $P\left(H\right)$; (iii) $s^i + m - c^i + z^i < -S^i \leq s^i + m - c^i$. In this case, which occurs for non-voters under *Stigma*, $P\left(SV\right)_{FE} = 0 \leq P\left(SV\right)_F = P\left(H\right)_F$; (iv) $s^i + m - c^i < -S^i \leq s^i + m - c^i + z^i$. In this case, which occurs for voters under *Pride*, $P\left(SV\right)_F = 0 \leq P\left(SV\right)_{FE} = P\left(H\right)_{FE}$. Under *Pride*, cases (i), (ii), and (iv) apply and pairwise comparisons for all these cases show $P\left(SV\right)_{FE}^v \geq P\left(SV\right)_F^v$. Under *Stigma*, cases (i), (ii), and (iii) apply and pairwise comparisons for all these cases show $P\left(SV\right)_{FE}^{nv} \leq P\left(SV\right)_F^{nv}$.

Turning to $P\left(SV\right)_{OO}$ and $P\left(SV\right)_{OOE}$, consider that, conditional on seeing the flyer, any person who answers the door will complete the survey. (Otherwise, this person could have costlessly opted out.) Therefore, the results on $P\left(SV\right)_{OO}$ and $P\left(SV\right)_{OOE}$ follow directly from the results on $P\left(H\right)_{OOE}$ and $P\left(H\right)_{OO}$.

Proof of Proposition 3. A voter will lie if $s_N^v - L^v + I \geq s_V^v$ or $-\left(s_V^v - s_N^v\right) - L^v \geq -I$. Under the assumption $s_V^v - s_N^v > 0$ and given $L \geq 0$, the left-hand side in the second expression is always negative; hence, a voter will never lie with no inducement ($I = 0$). And increase in I makes it more likely that the expression will be satisfied and thus (weakly) increases lying.

We consider then a non-voter. The lying condition for non-voters is $s_V^{nv} - L^{nv} \geq s_N^{nv} + I$ or $\left(s_V^{nv} - s_N^{nv}\right) - L^{nv} \geq I$. The left-hand side can be positive or negative depending on whether the net signalling utility or the lying cost is larger; hence, non-voters may lie even absent incentives I. Increased incentives I make it less likely that the inequality will be satisfied and hence (weakly) reduce lying.

Proof of Proposition 4. Individuals vote if the net expected utility in (1) is positive. Remembering that H is the c.d.f of $-(pV + g - c)$, we can rewrite the probability of voting as $H[N[\max(s_V, s_N - L) - \max(s_N, s_V - L)]]$. Under the assumptions $s_V - s_N > 0$ and $L > 0$, it follows that $\max(s_V, s_N - L) = s_V$ and that $s_V > \max(s_N, s_V - L)$. Hence, the term in square brackets is positive and the conclusion follows.

B Appendix B - Experiment Implementation

Each flyer distributor's participation in the study followed two steps: (1) an invitation to work as a paid volunteer for the research center and (2) participation as a distributor of flyers in the door-to-door campaign. Each surveyor's participation in the study typically followed four steps: (1) an invitation to work as a paid volunteer for the research center, (2) an in-person interview, (3) a training session, and (4) participation as a surveyor in the door-to-door campaign.

C Appendix C - Estimation Appendix

The classical minimum-distance estimator chooses the parameters $\hat{\xi}$ that minimize the distance given by $(m(\xi) - \hat{m})' W (m(\xi) - \hat{m})$. As a weighting matrix W, we use the diagonal of the inverse of the variance-covariance matrix. Hence, the estimator minimizes the sum of squared distances, weighted by the inverse of the variance of each moment. (Given the large number of moments, weighting the estimates by the inverse of the full variance-covariance matrix is problematic computationally. As a robustness check, we also use an identity weighting matrix, as reported in Appendix Table 2.)

To compute the theoretical moments $m(\xi)$, we derive closed-form expressions for the moments when possible (for example, the probability of being at home and completing the survey in the no flyer, survey flyer, and survey opt out treatments). When this is not possible, we use two types of numerical integration algorithms to calculate the moments. When integrating over a single random variable such as $z^v = \max\{s_V^v, s_N^v - L^v\}$ or $z^{nv} = \max\{s_N^{nv}, s_V^{nv} - L^{nv}\}$, we first derive expressions for the probability density function of z^i and then use highly accurate numerical integration algorithms based on adaptive Simpson quadrature, pre-implemented in Matlab as the *quadl* and *quadgk* routines. When double integrating over the distributions of two random variables with dependent limits of integration - the survey utility s and the social-image taste z^i - we use a 12-point non-adaptive Gaussian quadrature method for computational efficiency.

The empirical moments $\hat{m}$ are estimated in a first-stage model using the same controls as in the main regressions, and are listed in Appendix Table 1. In particular, all the moments other than the lying moments are calculated conditional on fixed effects for surveyor, day-town, and hour-of-day. We run OLS regressions with the relevant dependent variable (such as answering the door or completing the survey), treatment indicators for each of the relevant treatments, interacted with voters and non-voters indicators, as well as the demeaned fixed effects indicated above. (That is, we assume that the fixed effects have the same impact on voters and non-voters). We estimate these models jointly on the entire sample of voters and non-voters. For the probability of lying moments, no fixed effects are used, given the smaller sample of survey respondents.

Under standard conditions, the minimum-distance estimator using weighting matrix W achieves asymptotic normality, with estimated variance $(\hat{G}'W\hat{G})^{-1}(\hat{G}'W\hat{\Lambda}W\hat{G})(\hat{G}'W\hat{G})^{-1}/N$, where $\hat{G} \equiv N^{-1}\sum_{i=1}^{N} \nabla_\xi m_i(\hat{\xi})$ and $\hat{\Lambda} \equiv Var[m(\hat{\xi})]$ (Wooldridge, 2002). We calculate $\nabla_\xi m(\hat{\xi})$ numerically in Matlab using an adaptive finite difference algorithm.

To calculate the minimum distance estimate, we employ a common sequential quadratic programming algorithm (Powell, 1983) implemented in Matlab as the fmincon routine. We impose the following constraints: $\mu_j^i \in [-999, 999]$ for $i \in \{v, nv\}$ and $j \in \{V, N\}$ (finite social-image utilities), $\sigma_{SI}^i > 0$ (positive standard deviation of social-image utilities), $L \geq 0$ (non-negative lying costs), $S^s \geq 0$ (social pressure non-negative), $\sigma^s > 0$ (positive standard deviation of altruism), $h_0, r \in [0, 1]$ (probabilities between zero and one), $\eta \in [0, 99]$ (finite elasticity of home presence) and $v_s \in [0, 999]$ (finite and non-negative value of time). We begin each run of the optimization routine by randomly choosing a starting point, drawn from a uniform distribution over the permitted parameter space. The algorithm determines successive search directions by solving a quadratic programming sub-problem based on an approximation of the Lagrangian of the optimization problem. To avoid selecting local minima, we choose the run with the minimum squared distance of 120 runs. For the results presented here, the best estimate is achieved in about 15-40 percent of all runs, depending on the specification.

In the benchmark specification, the estimates for two closely related auxiliary parameters - the elasticity of home presence η and the social pressure cost S_s - are worth discussing in more detail. First, note that social pressure is largely identified by the fact that those who see the survey flyer and intend to say no to the surveyor optimally choose a home presence probability $h^* = h_0 - \eta S_s$ (barring corner solutions). Taking into account that only a share r sees the flyer, the observed probability of home presence under a flyer, assuming that all respondents intend to turn down the survey and again ignoring corner solutions, is $h_0 - r\eta S_s$. Of this expression, h_0 is precisely estimated from the share answering the door in the no flyer treatment, and r from the share checking the opt out box. Thus, the amount of sorting out between the no flyer and survey flyer treatments helps identify ηS_s. To disentangle these terms, the elasticity η is identified from the sorting in and out of the home as a function of the survey payment and duration. Interestingly, the benchmark estimate for ηS (Table 3) is similar for voters and non-voters, but the estimates of the individual parameters are very different. For voters, η^v is reasonably small (0.13, se = 0.09), and the social pressure cost sizeable, if smaller than in DellaVigna, List, and Malmendier (2012): S_s^v (\$1.76, se = 1.20). For non-voters, however, η^{nv} is extremely large and very imprecisely estimated, with then a tiny social pressure estimate. For such large value of η, non-voters are in practice always at corner solutions, and the exact value of η is not pinned down, given that any large η gives essentially the same solution. Notice that when we impose a fixed, small value of η (Column 6 in Appendix Table 2), the fit is substantially worse, but ultimately the estimated value of voting is about the same. The parameters η and S_s do not play a critical role for the estimate of the value of voting.

D Appendix D - Predicting the GOTV Effect

The turnout condition (1) is $N\Phi(s_V - s_N, L) + \varepsilon \geq 0$, where N is the number of times he expects to be asked whether he voted, $\Phi(s_V - s_N, L)$ is the social image motivation to vote, and ε indicates the remaining net benefits of voting. Conditional on a lying cost L, we estimated the distribution of the social image value of voting, Φ and we elicited the number of times asked N, but we do not have estimates for the distribution of ε. To calibrate what share of individuals will have their voter turnout decision changed by the GOTV treatment, we thus make assumptions about the distribution of ε.

We assume that the GOTV treatment flyer is perceived by the recipients as implying a probability h_0 of being asked *one more time* whether they voted, and thus an increase of $h_0 \cdot \Phi(s_V - s_N, L)$ in the expected social image value of voting. We pin down the mean of ε by requiring that the share of individuals with $N\Phi(s_V - s_N, L) + \varepsilon \geq 0$ matches the turnout in the GOTV control group on average. For various exogenous values of σ_ε, the standard deviation of ε, we can then simulate the effect of the GOTV treatment (i.e. the effect of increasing the expected utility of voting by $h_0 \cdot \Phi(s_V - s_N, L)$) for any given lying cost L. The detailed assumptions are as follows:

1. The GOTV flyer is seen with probability $r = 1$. This would tend to overstate the GOTV effect if some households do not observe or pay attention to the contents of the flyer, or do not believe it.

2. The household expects to be home during the future visit (with unspecified date) of the surveyor with the same baseline probability of being home during the survey experiment, h_0.

3. The social image parameters are estimated for households with all voters in 2010 ('Voters') and households with no voters ('Non-Voters'). We need to assign a voting probability to each group and approportion them in the GOTV sample. To do the first, we assume that each of these groups votes with an average probability. That is, the individuals in the 'Voters' households will have a higher voting probability than the individuals in the 'Non-Voters' household, but will not vote with probability 1 (or else the GOTV exercise is meaningless). For each of the two groups, we use the voting history to compute the predicted probability of voting based on past turnout, as we do in Section 6. For 'Voters', at the household level this implies a predicted probability of voting in the 2010 election of 0.71, thus $EVOTE^v = .71$. For 'Non-Voters', the implied probability of voting is 0.30, thus $EVOTE^{nv} = .3$.

4. The underlying share of 'Voters' p_v in the GOTV intervention then is calculated to match the observed turnout in the control group, 0.6. We thus have $p^v \cdot EVOTE^v + (1 - p^v) \cdot EVOTE^{nv} = p_v \cdot .71 + (1 - p_v)\, .71 = 0.6$, which implies $p^v = 0.75$.

5. We then make two assumptions meant to capture plausible lower bounds and higher bounds for the variance of other reasons to vote, ε. For the lower bound we assume that σ_ε equals the standard deviation of $s_v - s_N$, since it is plausible that the overall variance in all other reasons to vote across people is at least as large as the variance of the social image reason to vote. For the upper bound, we assume σ_e equal to 10 times the s.d. of $s_v - s_N$.

The procedure is then as follows:

1. Simulate $n = 1{,}000{,}000$ draws of s_V and s_N from the estimated distributions for 'Voters' and 'Non-Voters' from the benchmark specification in Table 3.

2. As a function of (assumed) lying cost L, calculate the social image motivation to vote according to Equation 2 for each of the simulated individuals.

3. Fix the mean of ε to match to turnout in the respective group. That is, for the 'Voters', calculate the $(1-EVOTE^v)^{th}$ percentile of the social image value of voting distribution, denoted by $\Phi^v_{m\arg inal}$, and set $\mu^v_\varepsilon = -\Phi^v_{m\arg inal}$. This ensures that we match the average turnout among 'Voters'. Then, we repeat this process for 'Non-Voters', using $EVOTE^{nv}$ and $\Phi^{nv}_{m\arg inal}$ to calculate μ^{nv}_ε.

4. For the assumed standard deviations of ε^v and ε^{nv}, simulate the draws of ε given the calculated μ^v_ε and μ^{nv}_ε from Step 3. For each individual, calculate their utility of voting for the N times asked (where N is a reported measure in our survey), $N\Phi(s_V - s_N, L)+\varepsilon$ and calculate the share with positive utility from voting. This is the control group GOTV turnout for each type. Average across the types using the share of 'Voters' (p^v) and 'Non-Voters' $(1 - p^v)$ to obtain the turnout in GOTV control group, $\tau_{control}$.

5. Calculate each individual's utility from voting in the treatment group, $(N+h_0)\cdot\Phi(s_V - s_N, L)+\varepsilon$ and the implied share with positive utility from voting in the treatment group. Average across the types to calculate the total GOTV treatment group turnout, τ_{GOTV}. The predicted GOTV effect is $\tau_{GOTV} - \tau_{control}$.

E Appendix E - Estimation of Lying Cost in Laboratory Experiment

Erat and Gneezy (2012) study lying behavior by conducting a sender-receiver game in the lab with 517 subjects. The game provides incentives for the "sender" to lie to the "receiver", for either altruistic or selfish motives. First, the sender is informed about the true outcome from rolling a six-sided die. She is then asked to send a cheap-talk signal of the outcome to the receiver. Next, the receiver chooses one of the six possible outcomes and, if this choice matches the state, payoff bundle A is implemented; otherwise, payoff bundle B is implemented. Importantly, the sender knows the payoffs A and B, while the receiver does not. The payoffs are varied to examine how lying by the sender depends on whether the lie is likely to help the receiver at a cost to the sender (an altruistic lie), help both the sender and receiver (a pareto lie), or help the sender at the cost of the receiver (a selfish or spiteful lie).

The payoffs for lying and truth-telling in each of five decisions are listed in Online Appendix Table 4, with the sender's payoff listed first. Thus, in Decision 1, lying results in a payoff of (19,30) - \$19 to the sender and \$30 to the receiver (assuming that the receiver chooses the signaled number). We assume a model of simple altruism with lying costs and model the sender as maximizing the utility function:

$$\max_{\{A,B\}} U = \{s_A + \alpha r_A, s_B + \alpha r_B - L + \varepsilon\}$$

where s_i is the sender's monetary payoff in outcome $i \in \{A, B\}$, r_i is the receiver's payoff, α is the sender's altruism towards the receiver, L is the psychological cost of lying and ε is a mean-zero utility shock to payoff bundle B (or equivalently, to payoff bundle A). To estimate the model, we impose the following assumptions: Lying cost L and altruism α are both assumed to be identical across individuals. The utility shock ε is distributed normally with mean zero and standard deviation σ_ε. We also assume that the receiver always follows the sender signal.

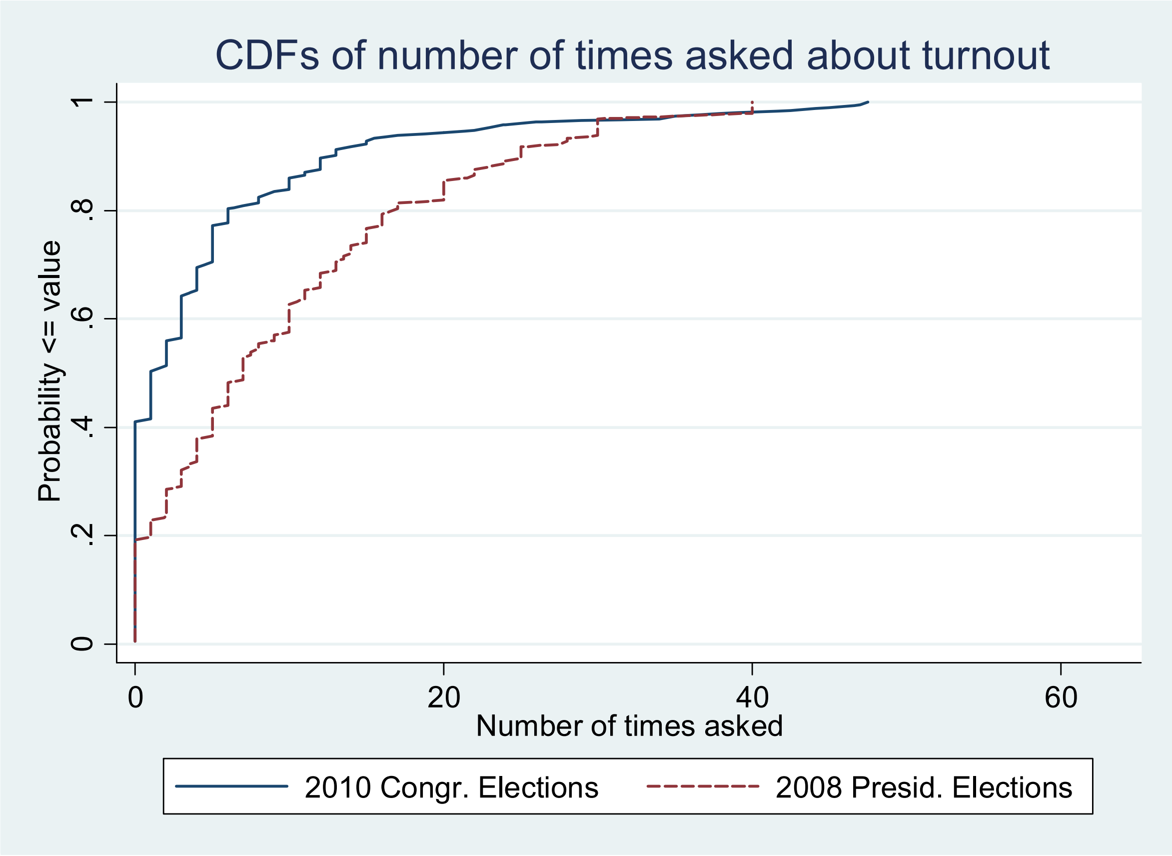
We estimate the model using a classical minimum distance estimator, with the shares lying in each decision as the five moments. The moments are weighted by the inverse of the variance of each moment. The intuition for the identification is straightforward. Conditional on altruism, the response of lying rates to the sender and receiver's monetary payoffs from lying identifies the lying cost as well as the variance of the error term.

The results suggests a substantial cost of lying, $L =$ \$7.0 (se \$1.4). The estimated lying cost is consistent with the reduced form observation that a third to a half of subjects choose not to lie even when the private gain from doing is \$10 (Decisions 3 and 5). The estimated altruism is $\alpha = 0.29$ (se 0.17) – senders value a dollar to the receiver as much as 29 cents to themselves. Finally, the standard deviation of the error term is $\sigma_\varepsilon =$ \$18.6 (se \$4.0). This heterogeneity is consistent with the fact that increasing the private incentive to lie from \$1 to \$10 increases lying by only 16 percentage points (Decision 2 vs. 3), suggesting a relatively low local density. At these estimated parameter values the fit of the moments is good, as Online Appendix Table 4 shows.

Extrapolated to the setting of our field experiment, this mean lying cost would imply a substantial social-image motivation for voting. In the benchmark specification, a lying cost of \$7 implies a social-image value of voting in congressional elections of \$16.9 for voters and \$18.8 for non-voters. Of course, we must be cautious in translating the lying cost estimated in this experiment to that in our survey experiments. One difference is that in our setting, the surveyor does not actually know if the respondent is lying (since our surveyors were blinded to the true voting status of the respondents and since the respondents likely are unaware that we know their voting status). In Erat and Gneezy (2012), in contrast, the sender knows that her lying or truth-telling is observed by the experimenter. In addition, the sample in Erat and Gneezy (2012) consists of undergraduate students, while our sample consists of adult voters and non-voters in Chicago suburbs.

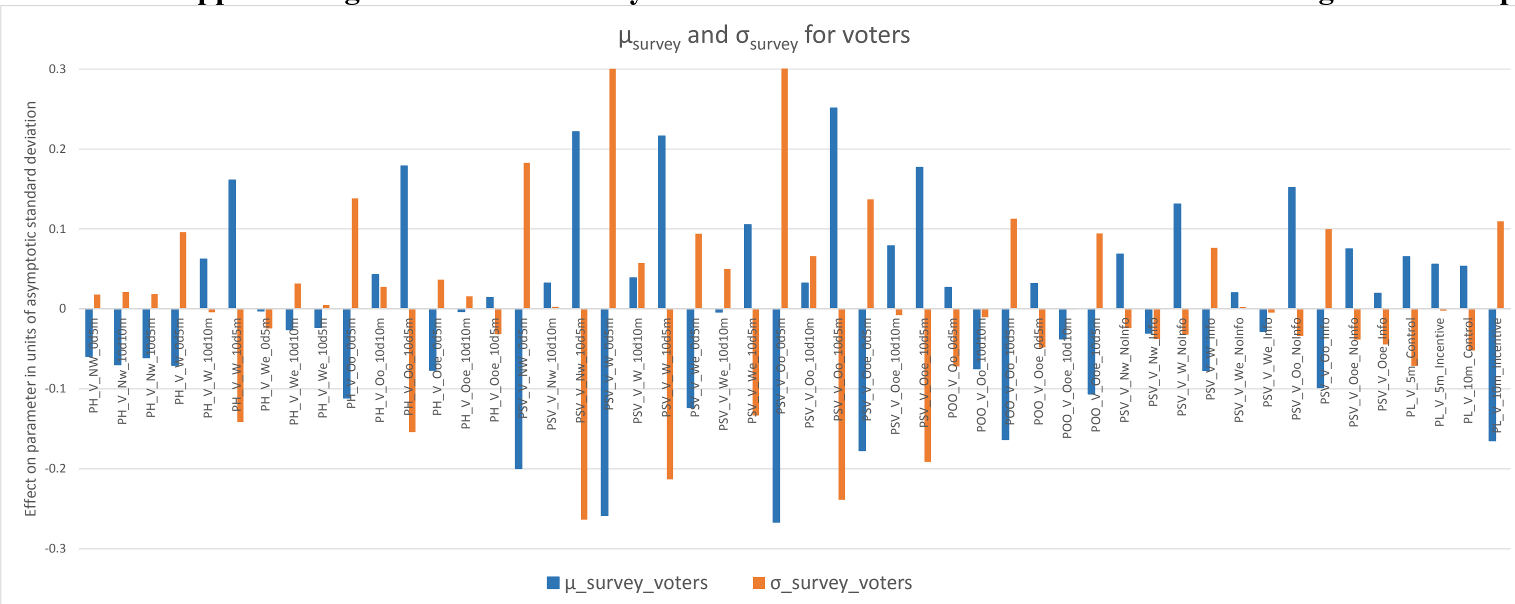
<u>For Online Publication Only</u>

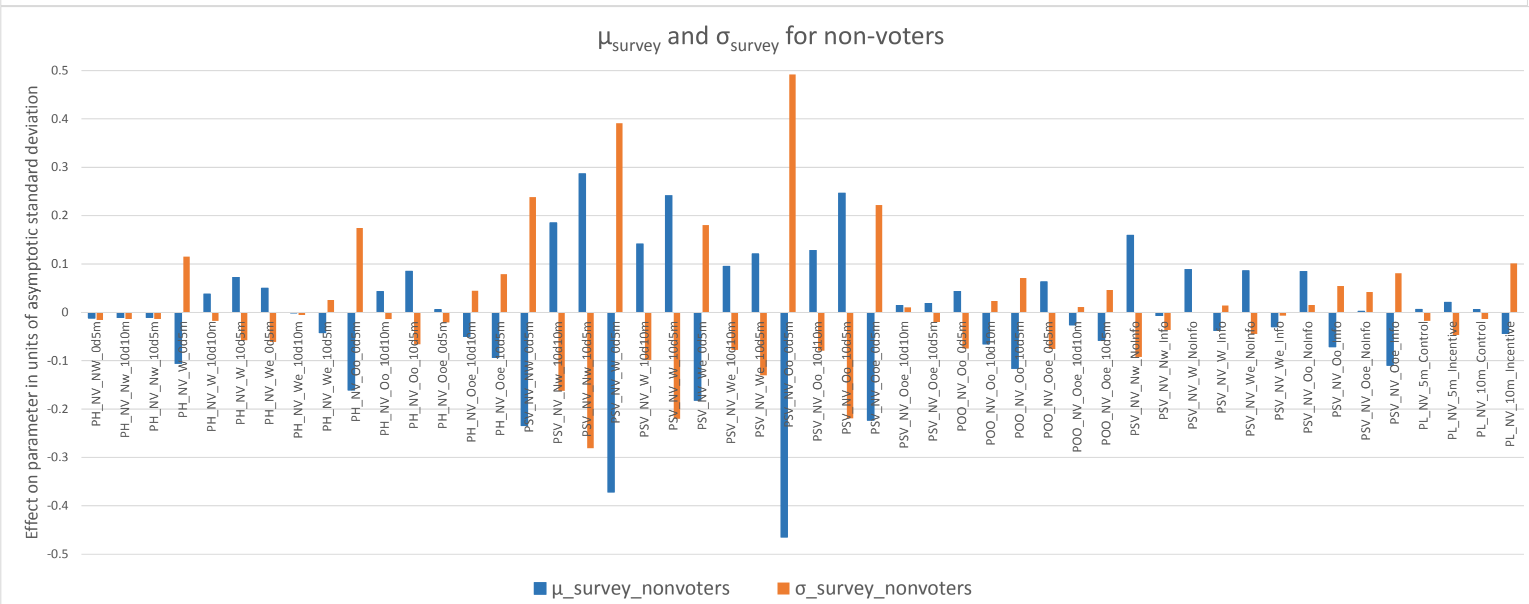
Online Appendix Figure 1. Number of Times Asked about Voting



Note: Online Appendix Figure 1 plots the cumulative distribution function of the self-reported number of times asked among the respondents to the 2011 door-to-door survey. The continuous line refers to the 2010 Congressional election, and the dotted line refers to the 2008 Presidential election.

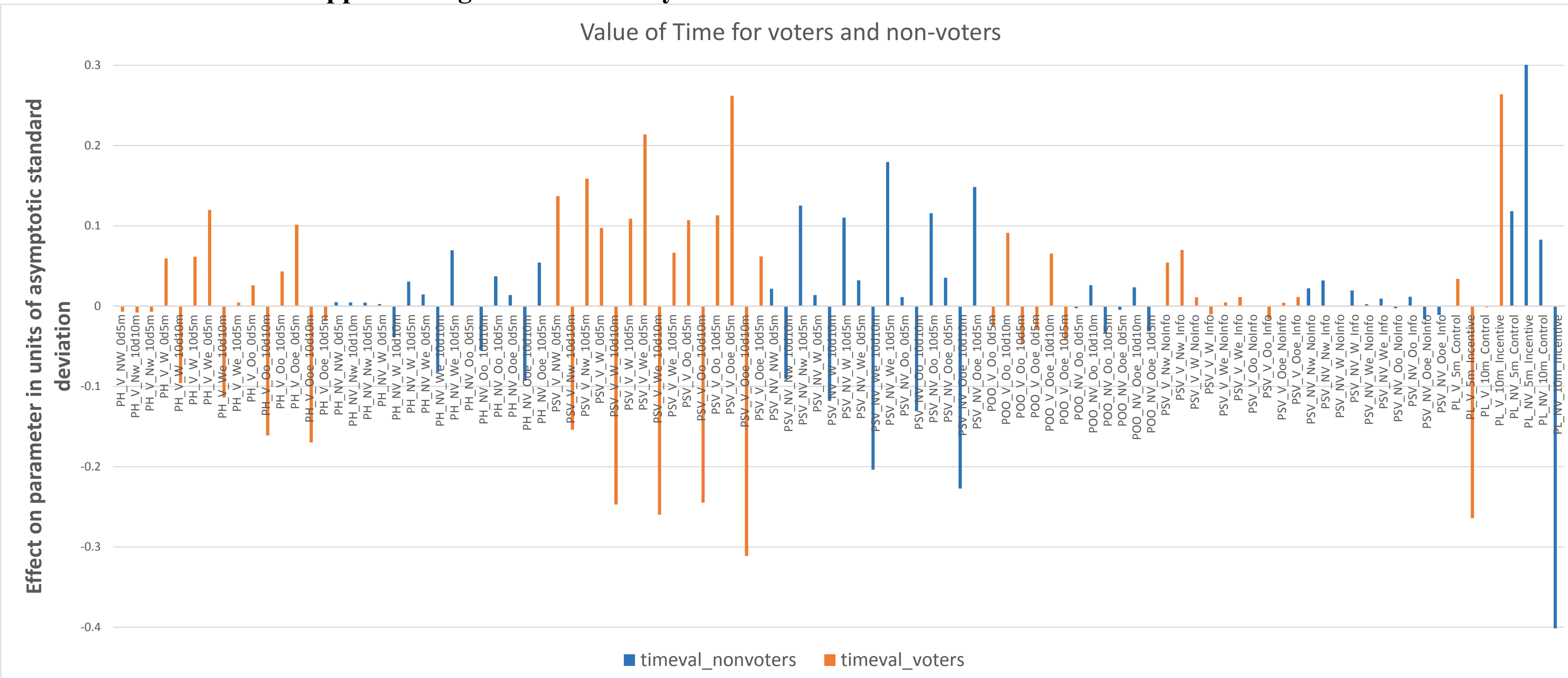
Online Appendix Figures 2a-b. Sensitivity of Estimated Parameters to the Moments: Willingness to Complete a Survey





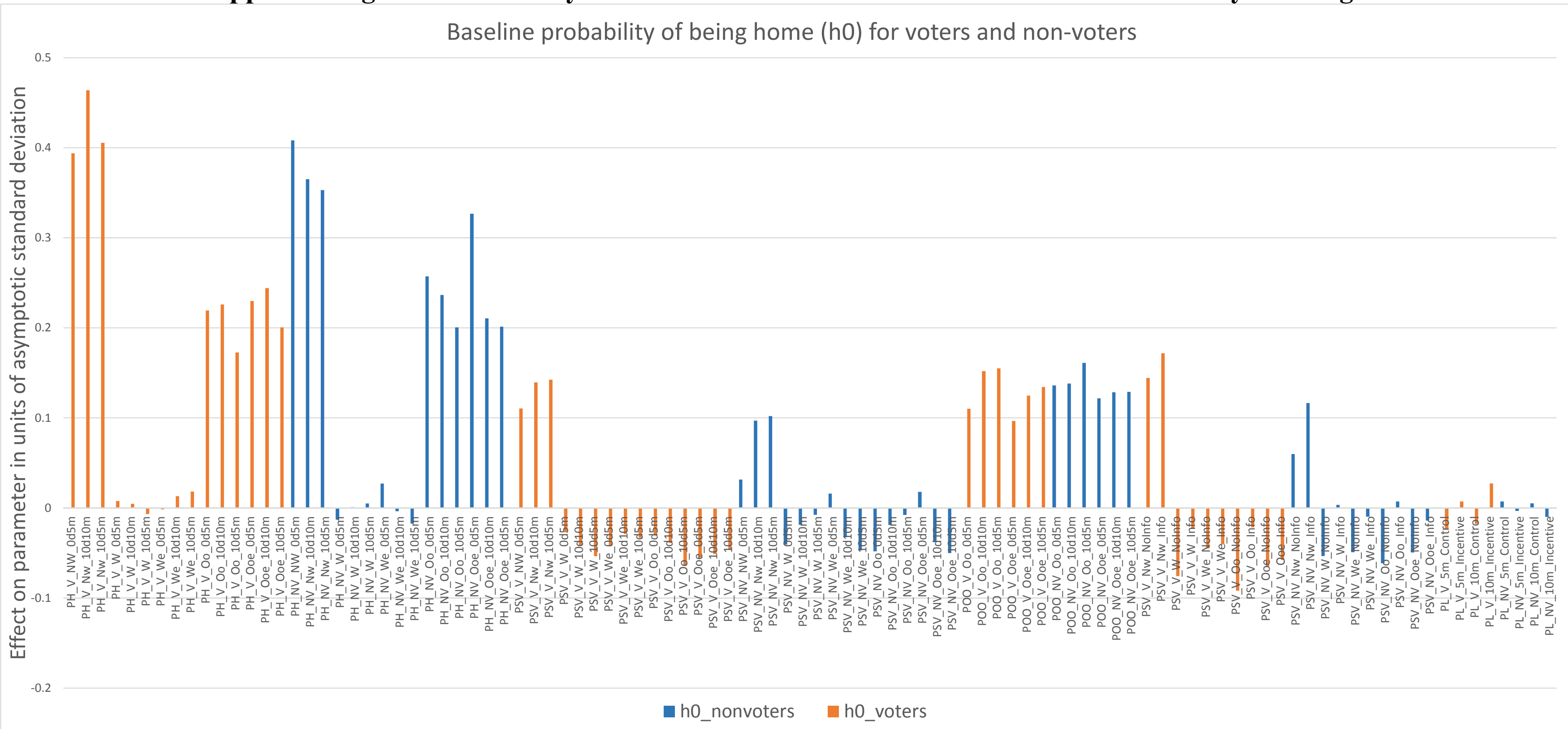
Note: Online Appendix Figures 2a and 2b present the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). The plotted lines indicate the local sensitivity of the given parameter estimate to an individual moment. A positive bar indicates that a (local) increase in the moment would increase the estimated value of the parameter. Higher bars indicate more influential moments for the identification of the parameter. For each moment, we plot the influence estimate for the mean value of completing a 10-minute survey (in blue) and the standard deviation of this willingness to complete a survey (in orange) respectively for voters (Online Appendix figure 2a) and for non-voters (Online Appendix Figure 2b).

Online Appendix Figure 3. Sensitivity of Estimated Parameters to the Moments: Value of time



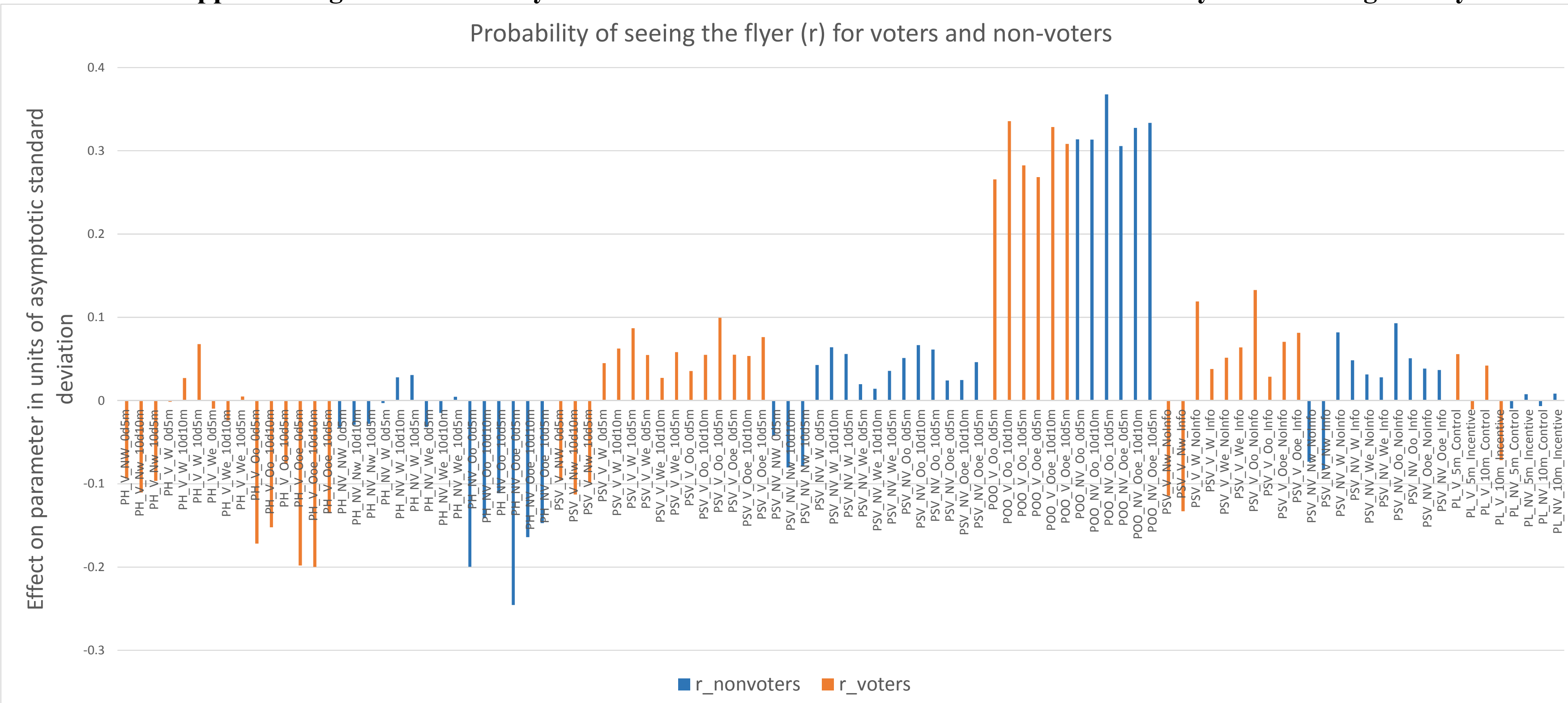
Note: Online Appendix Figure 3 presents the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). The plotted lines indicate the local sensitivity of the given parameter estimate to an individual moment. A positive bar indicates that a (local) increase in the moment would increase the estimated value of the parameter. Higher bars indicate more influential moments for the identification of the parameter. For each moment, we plot the influence estimate for the value of time for voters (in blue) and for non-voters (in orange).

Online Appendix Figure 4. Sensitivity of Estimated Parameters to the Moments: Probability of Being at Home



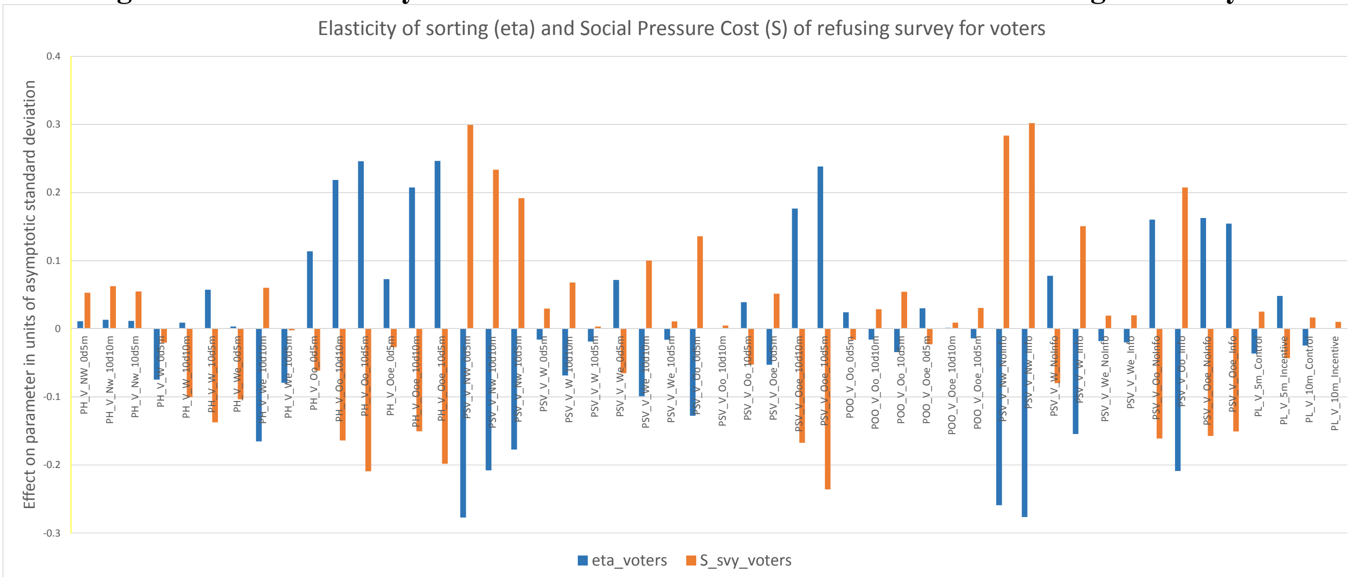
Note: Online Appendix Figure 4 presents the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). The plotted lines indicate the local sensitivity of the given parameter estimate to an individual moment. A positive bar indicates that a (local) increase in the moment would increase the estimated value of the parameter. Higher bars indicate more influential moments for the identification of the parameter. For each moment, we plot the influence estimate for the baseline probability of being at home for voters (in blue) and for non-voters (in orange).

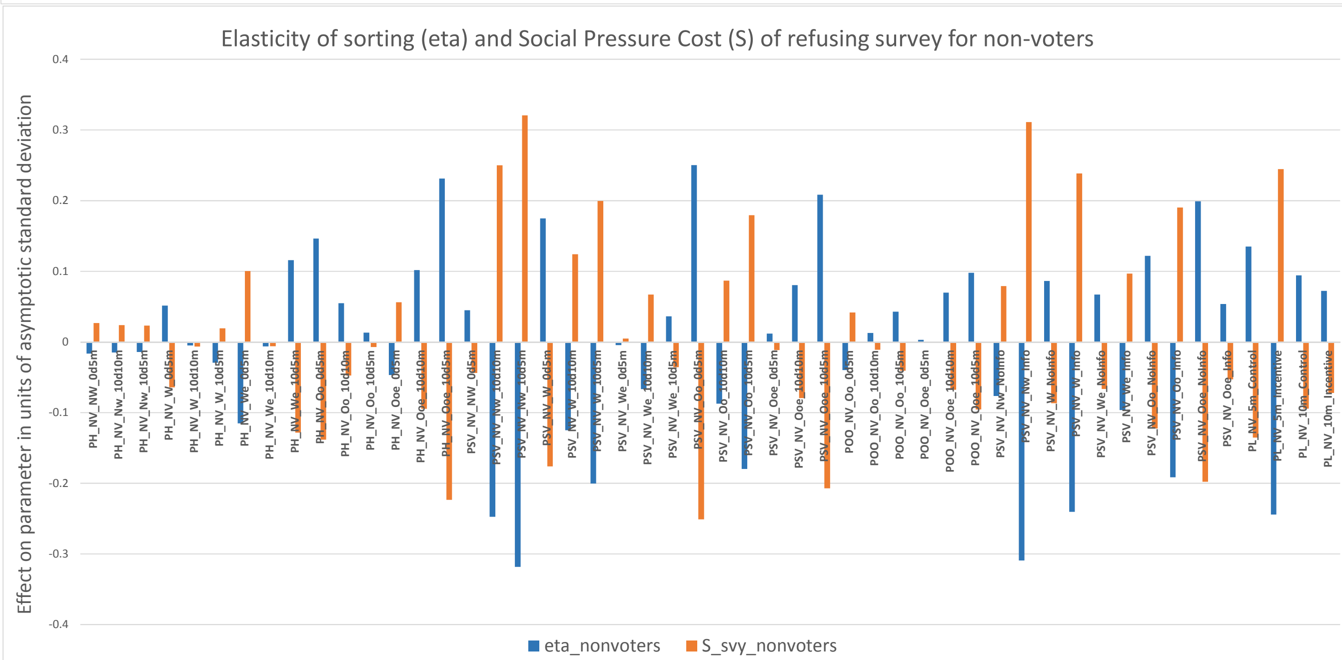
Online Appendix Figure 5. Sensitivity of Estimated Parameters to the Moments: Probability of Observing the Flyer



Note: Online Appendix Figure 5 presents the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). The plotted lines indicate the local sensitivity of the given parameter estimate to an individual moment. A positive bar indicates that a (local) increase in the moment would increase the estimated value of the parameter. Higher bars indicate more influential moments for the identification of the parameter. For each moment, we plot the influence estimate for the probability of observing the flyer for voters (in blue) and for non-voters (in orange).

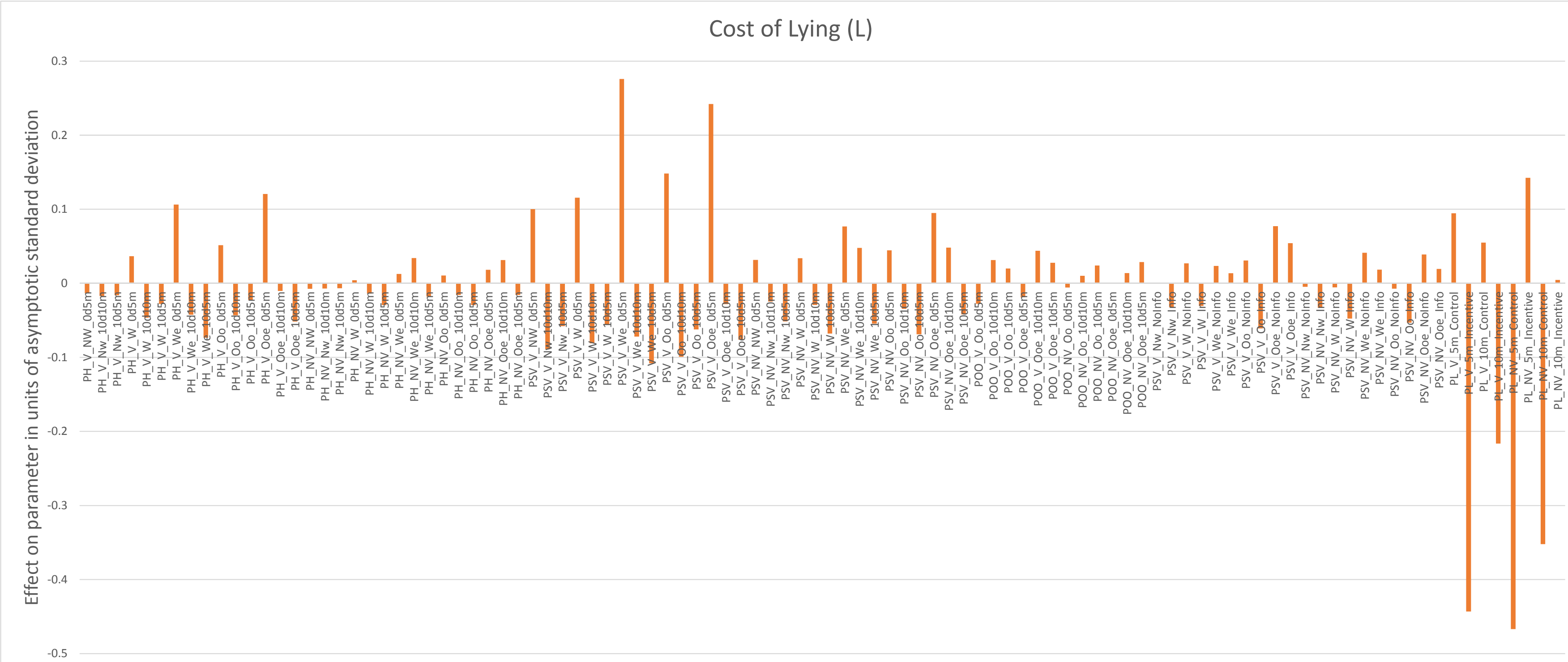
Online Appendix Figures 6a-b. Sensitivity of Estimated Parameters to the Moments: Sorting Elasticity and Social Pressure





Note: Online Appendix Figures 6a-b present the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). The plotted lines indicate the local sensitivity of the given parameter estimate to an individual moment. A positive bar indicates that a (local) increase in the moment would increase the estimated value of the parameter. Higher bars indicate more influential moments for the identification of the parameter. For each moment, we plot the influence estimate for the cost of sorting (in blue) and for the social pressure cost (in orange) first for voters (Online Appendix Figure 6a) and then for non-voters (Online Appendix Figure 6b).

Online Appendix Figure 7. Sensitivity of Estimated Parameters to the Moments: Lying Cost (Full Estimation)



Note: Online Appendix Figure 7 presents the sensitivity of the estimates to the individual moments following Gentzkow and Shapiro (2013). The plotted lines indicate the local sensitivity of the given parameter estimate to an individual moment. A positive bar indicates that a (local) increase in the moment would increase the estimated value of the parameter. Higher bars indicate more influential moments for the identification of the parameter. For each moment, we plot the influence estimate for the lying cost. These estimates are for the "full estimation" case, which requires that on this subsample voters and non-voters have the same key parameters, allowing for estimation of the lying cost.

Online Appendix Table 1. Survey Treatments, Robustness

Specification:	OLS Regressions							
Dependent Variable:	Indicator for Answering the Door				Indicator for Completing Survey			
Group:	Voters		Non-Voters		Voters		Non-Voters	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$10/10min Treatment	0.0364**	0.0314*	0.0243	0.0254	0.0132	0.0124	0.0231***	0.0266***
	(0.015)	(0.016)	(0.015)	(0.016)	(0.010)	(0.011)	(0.009)	(0.009)
$10/5min Treatment	0.0596***	0.0518***	0.0204	0.0196	0.0683***	0.0638***	0.0467***	0.0470***
	(0.017)	(0.018)	(0.015)	(0.017)	(0.013)	(0.014)	(0.009)	(0.010)
Simple Flyer Treatments	0.0128	0.0091	0.0286	0.0224	0.0960***	0.0948***	0.0496***	0.0510***
	(0.018)	(0.020)	(0.018)	(0.019)	(0.013)	(0.014)	(0.010)	(0.011)
Flyer Treatments with Opt-out	-0.0232	-0.0219	0.0052	0.0049	0.0695***	0.0731***	0.0325***	0.0349***
	(0.019)	(0.021)	(0.018)	(0.019)	(0.013)	(0.014)	(0.010)	(0.011)
Mention of Election in Flyer	-0.0143	-0.0206	-0.0278**	-0.0274*	-0.0194*	-0.0238**	-0.0238***	-0.0216**
	(0.013)	(0.014)	(0.014)	(0.015)	(0.011)	(0.012)	(0.008)	(0.009)
Voters Informed at Door of Election Topic					0.0001	-0.0018	0.0047	0.0085
					(0.009)	(0.010)	(0.008)	(0.008)
Omitted Treatment	No Flyer, $0/5min Treatment				No Flyer, $0/5min, Not Informed Treatment			
Fixed Effects for Solicitor, Date-Location, and Hour (Benchmark)	X		X		X		X	
Fixed Effects for Solicitor-Date-Location, and Hour		X		X		X		X
R2	0.0279	0.0629	0.0338	0.0765	0.0350	0.0650	0.0269	0.0734
N	6,873	6,873	6,324	6,324	6,873	6,873	6,324	6,324

Notes: Estimates for a linear probability model with standard errors, clustered by solicitor-date, in parentheses. The omitted treatment is the Baseline No-Flyer $0-5 minutes survey. The regressions include fixed effects for the solicitor, for the date-town combination, and for the hour of day in Columns 1 ,3, 5, 7. The regressions include in addition fixed effects for solicitor-date-town location in Columns 2, 4, 6, 8.

\ * significant at 10%; ** significant at 5%; *** significant at 1%

Online Appendix Table 2. Survey Treatments, By Time Period

Specification:	OLS Regressions											
Dependent Variable:	Indicator for Answering the Door				Indicator for Completing Survey				Indicator for Lie in Turnout Question			
Time Period:	Summer	Fall	Summer	Fall	Summer	Fall	Summer	Fall	Summer	Fall	Summer	Fall
Group:	Voters		Non-Voters		Voters		Non-Voters		Voters		Non-Voters	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$10/10min Treatment	0.0510***	0.0110	0.0072	0.0460**	0.0271**	-0.0096	0.0028	0.0498***				
	(0.018)	(0.026)	(0.020)	(0.023)	(0.014)	(0.016)	(0.012)	(0.012)				
$10/5min Treatment	0.0609***	0.0543*	0.0039	0.0434**	0.0654***	0.0700***	0.0432***	0.0534***				
	(0.020)	(0.030)	(0.022)	(0.021)	(0.016)	(0.021)	(0.013)	(0.013)				
Simple Flyer Treatments	0.0094	0.0167	-0.0007	0.0683***	0.0953***	0.0928***	0.0268*	0.0815***				
	(0.024)	(0.029)	(0.025)	(0.025)	(0.018)	(0.021)	(0.014)	(0.014)				
Flyer Treatments with Opt-out	-0.0204	-0.0299	-0.0181	0.0356	0.0766***	0.0545***	0.0208	0.0507***				
	(0.024)	(0.031)	(0.024)	(0.026)	(0.017)	(0.021)	(0.015)	(0.015)				
Mention of Election in Flyer	-0.0125	-0.0140	-0.0112	-0.0472**	-0.0080	-0.0331**	-0.0207*	-0.0273**				
	(0.018)	(0.019)	(0.018)	(0.021)	(0.015)	(0.016)	(0.011)	(0.012)				
Voters Informed at Door of Election Topic					0.0008	-0.0053	0.0064	0.0031				
					(0.012)	(0.013)	(0.011)	(0.011)				
Treatment with Incentive to Say that Did not Vote									0.0263	0.0162	-0.1502***	-0.0777
									(0.023)	(0.035)	(0.054)	(0.059)
Omitted Treatment	No Flyer, $0/5min Treatment				No Flyer, $0/5min, Not Informed Treatment				No Incentive to Lie			
Solicitor, Date-Location, Hour F.e.	X	X	X	X	X	X	X	X				
Date-Location F.e.									X	X	X	X
R2	0.0265	0.0325	0.0344	0.0341	0.0343	0.0423	0.0256	0.0353	0.0237	0.0745	0.0787	0.0670
N	4,245	2,628	3,459	2,865	4,245	2,628	3,459	2,865	718	418	343	252

Notes: Estimates for a linear probability model with standard errors, clustered by solicitor-date, in parentheses. The regressions include fixed effects for the solicitor, for the date-town combination, and for the hour of day in Columns 1-8 and fixed effects for date-location in Columns 9-12.
* significant at 10%; ** significant at 5%; *** significant at 1%

Online Appendix Table 3. Survey Treatments, By Political Registration

Specification:	OLS Regressions								
Dependent Variable:	Indicator for Answering the Door			Indicator for Completing Survey			Lie in Turnout Question		
Political Registration:	Republican	Democratic	Other	Republican	Democratic	Other	Republican	Democratic	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Voters									
$10/10min Treatment	0.0604**	0.0285	0.0450*	0.0272	0.0008	0.0273			
	(0.025)	(0.023)	(0.025)	(0.018)	(0.017)	(0.017)			
$10/5min Treatment	0.0544*	0.0550**	0.0887***	0.0827***	0.0612***	0.0677***			
	(0.028)	(0.025)	(0.029)	(0.021)	(0.019)	(0.021)			
Simple Flyer Treatments	-0.0169	0.0610**	-0.0250	0.0777***	0.1265***	0.0758***			
	(0.033)	(0.025)	(0.032)	(0.023)	(0.021)	(0.024)			
Flyer Treatments with Opt-out	-0.0769**	0.0322	-0.0593*	0.0687***	0.0801***	0.0564**			
	(0.035)	(0.025)	(0.035)	(0.024)	(0.021)	(0.023)			
Mention of Election in Flyer	0.0332	-0.0295	-0.0397	0.0046	-0.0290*	-0.0344*			
	(0.025)	(0.020)	(0.025)	(0.021)	(0.017)	(0.019)			
Voters Informed at Door of Election Topic				-0.0121	0.0242*	-0.0242			
				(0.017)	(0.014)	(0.017)			
Treatment with Incentive to Say that Did not Vote							0.0319	0.0275	0.0523
							(0.043)	(0.027)	(0.054)
Omitted Treatment	No Flyer, $0/5min Treatment			No Flyer, $0/5min, Not Informed Treatment					
Solicitor, Date-Location, Hour F.e.	X	X	X	X	X	X			
Date-Location F.e.							X	X	X
R2	0.0651	0.0476	0.0512	0.0701	0.0554	0.0658	0.1660	0.0590	0.0902
N	1,918	3,018	1,937	1,918	3,018	1,937	300	565	271
Panel B. Non-Voters									
$10/10min Treatment	0.0549	0.0315	0.0245	0.1233**	0.0071	0.0193**			
	(0.061)	(0.033)	(0.018)	(0.048)	(0.021)	(0.009)			
$10/5min Treatment	0.0078	0.0110	0.0241	0.0241	0.0461*	0.0440***			
	(0.072)	(0.039)	(0.017)	(0.051)	(0.026)	(0.010)			
Simple Flyer Treatments	0.0811	0.0463	0.0225	0.0916	0.0448	0.0505***			
	(0.093)	(0.047)	(0.020)	(0.065)	(0.027)	(0.011)			
Flyer Treatments with Opt-out	-0.0025	0.0141	0.0008	0.0560	0.0385	0.0292**			
	(0.091)	(0.044)	(0.019)	(0.065)	(0.028)	(0.011)			
Mention of Election in Flyer	-0.0832	-0.0433	-0.0181	-0.1148***	-0.0114	-0.0215**			
	(0.069)	(0.034)	(0.015)	(0.043)	(0.022)	(0.009)			
Voters Informed at Door of Election Topic				-0.0070	0.0046	0.0031			
				(0.043)	(0.019)	(0.009)			
Treatment with Incentive to Say that Did not Vote							-0.0000	-0.1998*	-0.0970**
							(0.265)	(0.112)	(0.045)
Omitted Treatment	No Flyer, $0/5min Treatment			No Flyer, $0/5min, Not Informed Treatment					
Solicitor, Date-Location, Hour F.e.	X	X	X	X	X	X	X	X	X
Date-Location F.e.							X	X	X
R2	0.2710	0.0816	0.0381	0.2945	0.0706	0.0309	0.3762	0.2994	0.0986
N	351	1,179	4,794	351	1,179	4,794	42	126	429

Notes: Estimates for a linear probability model with standard errors, clustered by solicitor-date, in parentheses. The regressions include fixed effects for the solicitor, for the date-town combination, and for the hour of day in Columns 1-8 and fixed effects for date-location in Columns 9-12.
* significant at 10%; ** significant at 5%; *** significant at 1%

Online Appendix Table 4. Moments and Estimates on Erat and Gneezy (2012)

Decision Number:	Payoffs of A (Truth)	Payoffs of B (Lie)	Fraction Lying (Empirical)	Fraction Lying (At Estimated Parameters)
1	(20, 20)	(19, 30)	33/101 (33%)	39%
2	(20, 20)	(21, 30)	49/101 (49%)	43%
3	(20, 20)	(30, 30)	66/102 (65%)	62%
4	(20, 20)	(21, 15)	38/104 (37%)	34%
5	(20, 20)	(30, 20)	57/109 (52%)	56%

Parameter Estimates:	Lying Cost	Altruism Coefficient	S.D. of error term
	7.0 (1.4)***	0.29 (0.17)*	18.6 (4.0)***

Notes: Estimates from minimum-distance estimator using the 5 moments shows above and weights given by the inverse of the variance of each moment.
* significant at 10%; ** significant at 5%; *** significant at 1%