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
Harvard University and NBER

This version: May 13, 2016

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 'to tell others' of about \$15, contributing 2 percentage points to turnout. Lastly, we evaluate a get-out-the-vote intervention in which we tell potential voters that we will ask if they voted.

*We thank Nageeb Ali, Leonardo Bursztyn, David Card, Raj Chetty, Justin Fox, Andrew Gelman, Alan Gerber, Yosh Halberstam, Mitch Hoffman, Greg Huber, Rebecca Morton, 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, the 2013 SITE Psychology and Economics conference and the 2015 NYU-Abu Dhabi conference in Behavioral Political Economy for helpful comments. We also thank Alec Brandon, Tristan Gagnon-Bartsch, Sheng Li, David Novgorodsky, Jessica Shui, Alex Steiny, 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 such intervention to date, a letter revealing the recipient's and their neighbor's turnout record increases turnout by 8 percentage points (Gerber, Green and Larimer, 2008). This remarkable effect likely reflects social-image concerns: the explicit threat to make one's 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 (in part) 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 established fact that 25 to 50 percent of non-voters lie when asked about their past turnout (Silver, Anderson, and Abramson, 1986; Belli et al., 1999).[1]

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 due to this social image motivation. We consider this 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 visit their home to ask them to complete a survey. In a second group, the flyer specifies that the survey will be about "*your voter participation in the 2010 congressional election.*" Differences in the share of households opening the door and completing the survey between the first and the second group reflect the anticipated value of being asked about voting. An

[1] Memory failures or inaccurate voting records do not appear to explain the discrepancy (Duff et al., 2007).

increase in the participation of voting households would indicate the pride of saying that one voted. A decrease among non-voting households would indicate shame from admitting that one did not vote, or a cost of lying and claiming to have voted.[2]

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.[3]

These results may depend on the particular election. 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 '*to tell others*,' 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.

Thus, in a third set of crossed treatments we randomize incentives to provide a different response to the turnout question. 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 without such incentives, 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 to other settings where responses could have social image or signaling 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-

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

[3]We also cross-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 did not have a significant effect on survey 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 simulated minimum-distance estimator. The estimation accounts for the fact that the social-image variables affect not just the predicted response to the experimental treatments, but also whether an individual would choose to vote in the first place. Individuals select into being voters or non-voters as a function of the social-image values and of their other reasons to vote, such as expressive voting or voting costs. Thus, a voter is more likely to care about social image than a non-voter, and also more likely to have, say, low voting costs.

The benchmark estimates provide evidence of significant social-image value in voting. We estimate that individuals assign on average a \$7 value to being seen as a voter rather than a non-voter, when asked once by a surveyor. This social-image value does not come from pride in voting: in fact, individuals on average prefer not to be asked, even when they can say truthfully that they voted. Rather, they assign a quite negative value to admitting to not voting.

We estimate that individuals assign a disutility of \$7 to lying about voting to a surveyor. The combination of social-image utility and sizable lying costs implies that the anticipation of being asked provides a reason to vote.

To quantify the value of voting 'to tell others', we combine these estimates with survey-based evidence on the number of times that people report being asked whether they voted, 5 times on average for the 2010 congressional election. Altogether, we estimate a value of voting 'to tell others' of \$18 for voters and \$13 for non-voters.

There are caveats to this estimate. For example, this magnitude likely understates the value of voting 'to tell others', since it is based on being asked by a (previously unknown) surveyor. Even taking these caveats into account, our ability to assign a dollar value to voting through the design is a unique contribution to the literature. For example, 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 for the value of voting are up to a scaling for the voting cost, which is not identified; thus, they do not provide a monetary value of voting due to their model. Levine and Palfrey (2007) estimate a pivotal-voting model, but use laboratory elections where parameters can be controlled. The unique element in our design that makes the difference is that we use monetary inducements—variation in the value of the survey and incentive to lie about voting—to translate the findings into a monetary value of voting.

A second metric to evaluate the model of voting 'to tell others' is in terms of the extra turnout that it generates. The baseline turnout in our setting is 60 percent. How much would turnout change if people stopped asking others whether they voted, for example because of a shift in norms? What about if conversely the rate of asking doubled, perhaps because campaigns encourage such asking? We predict that eliminating asking about voting would lower turnout by 2 percentage points, while doubling the number of times asked would increase turnout by 2 points. While these impacts may seem small, consider the extraordinary amount of effort that campaigns put into get-out-the-vote efforts, with the average such letter yielding a turnout impact of 0.2 percentage points (Green, McGrath, and Aronow, 2013).

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.*'[4] 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, Ternovski, and Yoeli (2016), who also inform a treatment group that they may be called after the election about their voting behavior, and find a positive impact on turnout (of 0.2 percentage points). 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.[5]

Are the get-out-the-vote results consistent with the estimates of the value of voting? Using the model estimates, we predict that an announced visit to ask about voting would increase turnout by 0.3 percentage points, well within the point estimates of the estimated effects. Thus, the get-out-the-vote results are qualitatively consistent with the model, if imprecise.

Finally, we would like to mention some caveats and alternative interpretations. First, we address the important concern that the observed 'sorting out' among non-voters may reflect a dislike of talking about politics, rather than any stigma from not voting. When we allow for a different taste among voters and non-voters for talking about politics, we lose the ability to estimate one of the social-image parameters. But the net value of voting *'to tell others'* is

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

[5]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.

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.

Second, 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.

Third, in a series of robustness and sensitivity checks, we relax and vary numerous assumptions of the model, including modelling heterogeneity in lying cost, allowing for measurement error in the voting record, and omitting groups of moments. The estimated value of voting *'to tell others'* remains largely robust to these variations.

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.[6] Our main contribution is to provide an estimate of the value of voting and a welfare evaluation of a get-out-the-vote campaign, which is rare in the literature. We obtain these estimates by virtue of the design of the field experiment.

The paper also relates to the literature on social image. The theoretical papers provide a micro-foundation for social-image concerns as signaling models (Benabou and Tirole, 2006; Andreoni and Bernheim, 2009; 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 economy, including Shue and Luttmer (2009), Finan and Schechter (2012), Passarelli and Tabellini (forthcoming), 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

[6]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.

reduced-form results and structural estimates for the main experiment. Section 6 introduces the get-out-the-vote experiment and 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 ε 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 N times whether she voted, and has to decide whether to be truthful or to lie. Assume first that she has truly 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 signaling 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. 2016).

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 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.[7]

[7] The model assumes that the respondents do not have an option of refusing to answer the vote question. This seems justified by the data: out of 1,738 people that agreed to do the survey, the answer to the turnout question is missing in only 5 cases.

We stress four important assumptions. First, the variables s_V, s_N, and L could depend on a variety of factors, such as whether one's party won the election, or the closeness of the race. One could envision a model, for example of social signalling, which would provide a micro-foundation for the values of the variables.[8]

Second, we can relate this model to a more general model. Assume that the utility $U_{k,s}$ of being asked about voting depends on whether one voted or not ($k = V, N$), and on whether one tells the truth or lies ($s = T, L$). Our model can be mapped to such a model: $U_{V,T} = s_V$, $U_{V,L} = s_V - L$, $U_{N,T} = s_N$, and $U_{N,L} = s_N - L$. We impose the restriction that the lying cost in the two cases (V, N) is equal.

Third, we assume that the value of being asked increases linearly in the number of times asked N, an assumption that is untested. Alternatively, the value of being asked may be concave in the number of times asked, with later asks yielding less disutility, say, from lying about voting. If that is the case, we likely are underestimating the value of voting to tell others. By the time we ask voters about their past turnout, months after the election, respondents on average have already been asked multiple times. The value of our marginal ask, which we estimate with our treatments, would understate the value of being asked the previous times.

A final point is about heterogeneity. We assume that s_V, s_N, and ε are stochastic and heterogeneous. As we explain in Section 5, individuals become voters or non-voters following (1) depending on the draws of the variables, thus inducing systematic differences between voters and non-voters in these variables. In the benchmark model, for simplicity we assume that the lying cost L is instead deterministic and identical for voters and non-voters, though we relax this assumption in a robustness check. In another robustness check, we also estimate a model that allows for differences between voters and non-voters in the utility of talking about politics, another form of heterogeneity.

Returning to our model, 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 $N\Phi(s_V - s_N, L) + \varepsilon \geq 0$, where

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

As expression (2) shows, voting depends on the net social-image value $s_V - s_N$ and on

[8]Indeed we present evidence in our setting suggesting different values for registered Republicans and Democrats.

the cost of lying L. Figure 1 displays $\Phi(s_V - s_N, L)$ as a function of $s_V - s_N$ 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 $c(h)$, 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 $c(h) = (h - h_0)^2 / 2\eta$.

If the individual is at home at the time of the surveyor's visit and opens the door, 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. As in DellaVigna, List and Malmendier (2012), the respondent pays a social pressure utility 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 does not open the door when the surveyor visits. We 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 opening 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 $s_N - L + I \geq s_V$. In contrast, a non-voter is provided an incentive to tell the truth, and will do so if $s_N + I \geq s_V - L$. 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 opening the door, the respondent agrees to the survey if $s + m - c + z \geq -S$ assuming the respondent knows that the survey is about the election and if $s + m - c \geq -S$ otherwise. Working backwards, consider a respondent who sees a survey flyer (which does not mention the election questions). The decision problem of staying at home and opening the door (conditional on seeing a flyer) is $\max_{h\in[0,1]} h \max(s + m - c, -S) - (h - h_0)^2 / 2\eta$, leading to the solution $h^* = \max[\min[h_0 + \eta \max(s + m - c, -S), 1], 0]$. An increase in pay m or a decrease in the time cost c will increase the probability of opening the door and completing a survey. The parameter η determines the responsiveness of opening the door to incentives. Alternatively, for a respondent who sees the election flyer the solution is given by $h^* = \max[\min[h_0 + \eta \max(s + m - c + z, -S), 1], 0]$. If $z > 0$, the respondent will open the door with a weakly higher probability with the election flyer, compared to the survey flyer, and vice versa if $z < 0$.

Finally, for both the survey flyer and the election flyer, there is a variant with an opt-out box (denoted by OO and OOE, respectively) which makes avoidance of the surveyor easier. In terms of the model, this is equivalent to the agent being able to costlessly reduce the probability of being at home and opening the door to zero. Formally, $c(0) = 0$ and $c(h)$ is as above for $h > 0$.[9] The optimal probability of being at home and opening the door 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

[9] 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.

intervention (Proposition 4).[10]

Proposition 1. (Pride in Voting) *With Pride in Voting, the probability of opening the door $P(H)$ 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 opening the door $P(H)$ 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 $L > 0$.

3 Experimental Design

Logistics. We employed 50 surveyors and many flyer distributors, mostly undergraduates at the University of Chicago, who were paid $10.00 per hour. Most surveyors conducted surveys over multiple weekends.[11] 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 Appendix Figure 1.[12]

The unit of treatment assignment is a route, which consists of typically 13 households on a street, to be reached within a half-hour. On a day, a surveyor has a 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

[10] The proofs are in the Appendix.

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

[12] 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 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 households in the sample, 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[13]: 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 2). 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 2). 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 2). 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 2 displays flyers for 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*

[13] The ballot for the 2010 Congressional elections included, in addition to House and Senate races, ballots for governor and lieutenant governor, attorney general, secretary of state, comptroller and treasurer, and state senators. There are also Cook county-specific candidates and a proposed amendment to the Illinois constitution, and two town-specific referenda. The items on the ballot would largely be the same in the towns surveyed. Any differences, such as in the local referenda, would be controlled by town fixed effects (with the minor exception of towns spanning different congressional districts).

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 3 summarizes this first set of crossed treatments.

The fourth set of crossed treatments, summarized at the bottom of Figure 3, 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.[14]

Overall, we conducted 60 treatments: 5 flyer types crossed with 3 survey lengths, whether or not the survey content was announced at the door, and 2 types of lying incentives.

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. The complete survey transcript is in the Online Appendix.

[14]To keep a parallel structure, in the treatments with no incentive to lie about voting, there is instead an incentive to lie about the year of house purchase.

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).[15] The final sample includes 13,197 households.

4 Reduced-Form Estimates

Opening the door and survey completion. We present graphical evidence in Figure 4 on the share of households opening 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 opening 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 opening the door. The effect is similarly large for the share completing the survey, a 6 percentage points (45 percent) increase. The responsiveness to incentives of non-voters is smaller with regards to opening 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 5a 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 opening 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 of voters opening the door and in the share completing a survey when the flyer informs about the survey content compared to when it does not. 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 5b), 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 opening 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

[15] The rate at which the subjects are dropped is comparable across the different flyer treatments, but is higher in the no-flyer treatments (14% versus 8%). The reason is that households with a no-solicitor sign in the flyer treatments are excluded altogether from the sample when flyerers find the no-solicitor sign on their flyering visit; these houses are not visited the next day, to save time. This does not happen in the no-flyer treatment since there is no flyering visit. Thus, the no-fylering treatments include in the sample more no-solicitor households (previous to us dropping them). This being said, this difference plays a minimal role since the no-flyer treatments only help to identify auxiliary parameters. As we show in Column 5 of Online Appendix Table 5, the results are similar if we do not drop any observations.

in the opt-out treatments, with a 1.5 percentage point (15 percent) decrease in the share opening 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 6 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).[16] 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 7 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 opening 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 4, 5a and 5b, 6, and 7 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:

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

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×date level.[17]

[16] 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.

[17] For space reasons, the specification in Table 1 assumes an additive effect between the flyer treatments, the

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 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.[18] Finally, in Online Appendix Table 3 we present the results split by political registration, as in Figure 6.

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 location-day fixed effects η_i are included in regressions.[19]

In Table 2 and Figure 8, 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.[20]

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

payment and duration treatments and the door information treatments. The empirical moments used for the estimation, listed in Appendix Table 1, are more disaggregated.

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

[19] In Online Appendix Table 4, we show that the results are not sensitive to adding the full set of fixed effects.

[20] 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.

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 8 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.

Summary. To summarize, among voters we find little sorting on average into opening the door 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, a low social-image value of being a non-voter s_N^v , 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 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: $s_V^{nv} - s_N^{nv} = L^{nv}$. 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. The key variables of interest—the social image variables s_V and s_N and the lying cost L—affect not only the decision whether to answer the survey and the reporting of whether one voted, but also the turnout decision itself. For example, individuals with high social-image utility (high $s_V - s_N$) are more likely to state that they voted, even if they did not. These same individuals are also more likely to choose to vote in the first place, as equation (1) clarifies.

The model estimation procedure acknowledges this dual role of social image, first in determining the voting status, and then in affecting the behavior in the experiment. *Concretely*, we draw a simulated population of potential voters with values for social image, lying costs, and other parameters. Then, the individuals in the population choose to become voters or non-voters, depending on the draw of the variables. Accounting for the selection into voters and non-voters, we simulate the behavior of both groups in the various experimental conditions: the share opening the door and doing the survey in a particular condition, the share lying conditional on answering the survey, etc. Hence, the distribution of the key variables of interest, such as social image, will differ for voters and non-voters in ways predicted by the selection model, even though they are initially drawn from the same population distribution.

An alternative approach to estimation, which we adopted in earlier versions, is to assume different distributions of the variables for voters and non-voters *without* modelling the selection into voting. This second approach has two disadvantages. First, this approach ignores a key prediction of the model, that voters and non-voters will differ in a particular way, for example with voters displaying higher social-image concerns. Second, given that it does not model the selection into voting, there is less power to identify the parameters and it is not possible to pin down either the lying cost L or the other reasons to vote ε. Thus, we adopt the estimation with selection as the main set of estimates, but also report below the estimates of the model with exogenous assignment into voters and non-voters.

The key variables in the model of 'voting to tell others' are the social-image value of saying one voted, s_V, the social-image value of saying one did not vote, s_N, and the lying cost, L. 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 ($s_V < s_N$). The lying cost L for tractability reason is assumed to be deterministic, though we relax this assumption below in a robustness check.

Together with the residual value of voting ε, these variables determine the decision to vote according to equation (1). The residual value of voting ε is normally distributed with mean μ_ε and standard deviation σ_ε, and independently drawn from the signalling utility variable s_V and s_N. Thus, there are six parameters of interest: (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; (v) μ_ε, the mean of other reasons to vote, and (vi) σ_ε, the standard deviation of the other reasons to vote. These parameters suffice to determine the conditions for selection into voters and non-voters.

While these are the relevant parameters for the voting decision, additional auxiliary parameters affect the decision to open the door and answer the survey, and whether to lie about voting. We assume a utility s of completing an unpaid 10-minute survey, distributed normally with parameters μ_s and σ_s; the utility s is independent from s_V, s_N, and ε. We also assume a quadratic cost of changing plans to be at home, $c(h) = (h - h_0)^2 / 2\eta$ in case the person sees the flyer (which occurs with probability r). We thus have 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[21]; (vi) S_s, the social pressure cost associated with saying no to the survey request.

In our preferred specification, we allow voters to differ from non-voters in these auxiliary parameters.[22] In this specification, the total number of parameters is 20, with 6 key parameters

[21]Notice that the estimated value of time refers to minutes of doing a survey, which may differ from the value of time involved in deciding whether to vote.

[22]This can be thought of as an (unmodeled) correlation between the auxiliary variables and ε.

and 7*2 auxiliary parameters. In an alternative, more parsimonious specification we force the auxiliary parameters to be constant across voters and non-voters, for a total of 13 parameters.

To estimate the model, we use a simulated method of moments estimator. Denote by $m_N(\xi)$ the vector of simulated moments as a function of the parameters ξ, and by $\hat{m}$ the vector of observed moments. The estimator chooses the parameters $\hat{\xi}$ that minimize the distance $(m_N(\xi) - \hat{m})' W (m_N(\xi) - \hat{m})$. 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. For each run, we use a simulated population of at least 500,000 individuals. We discuss further details in Appendix B.

To list the moments $m_N(\xi)$, 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, $l \in \{NoInc, 8\min, \$5\}$. The moments $m_N(\xi)$ are: (i) the probability opening the door in treatments k, m, $P(H)^i_{k,m}$; (ii) the probability of completing the survey in 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: $P(H)^i_{a,k}$, (v) the probability of lying about past turnout conditional on completing the survey, given incentive l, $P(L)^i_l$, and (vi) the average turnout $P(V)$.[23] The empirical moments $\hat{m}$, 101 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.[24]

Identification. Regarding the main sources of identification, consider first the social-image and lying parameters, μ_V, μ_N, σ_{SI}, and L. 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 pin down the mean social-image utility $E^v[s_V]$, the expected value of s_V for individuals who select to become voters. 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, $E^{nv}[s_N]$, must approximately equal the utility from lying, $E^{nv}[s_V] - L$. A similar role is played by the difference in survey completion between the Information and No Information treatments.

Notice that, due to selection, the conditional expectations above do not equal the population

[23]The turnout moment is the turnout in the control group of the 2010 get-out-the-vote intervention, discussed in the next Section.

[24]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.

mean. For example, individuals with high social-image utility of saying one voted, s_V, are more likely to choose to be voters; thus, $E^v[s_V] > \mu_V > E^{nv}[s_V]$. Thus the sorting moments, which pin down the conditional expectations, do not suffice to identify μ_V, μ_N, and L. The missing element is the extent of selection into voters and non-voters with respect to the social-image variables, which determines deviations such as $E^v[s_V] - \mu_V$. We return to the selection below.

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 $\Pr(s_N < s_V - L < s_N + (8/60)\, v_s^{nv} | nv) = 0.12$.

As we mentioned above, a critical piece for identification is the extent of selection on social-image variables s_V and s_N, versus on the residual voting variable ε. This selection will depend crucially on the comparison between the variance σ_ε^2 in the residual reasons to vote, versus the variance σ_{SI}^2. If the variance σ_ε^2 of the residual reasons to vote ε is large relative to the variance of social image terms, σ_{SI}^2, the selection into voters and non-voters will mostly depend on other reasons to vote ε, as opposed to variation in social image. In this case, indeed $E^v[s_V] \approx E^{nv}[s_V] \approx \mu_V$ and thus the parameters μ_V, μ_N, and L are identified by the sorting moments alone. Conversely, consider the case in which the variance σ_ε^2 is small relative to the variance σ_{SI}^2 in social-image utility. In this case, there will be large differences in social image s_V and s_N between voters and non-voters. Thus, the estimate of σ_ε^2 plays a critical role in tying together the parameters estimated for voters and non-voters.

This leads us to the identification of the parameters for residual voting ε, μ_ε and σ_ε. A critical moment is the baseline turnout rate, which is precisely measured at 0.60 in this population. For any given value of σ_ε, the value of μ_ε is identified by this moment (taking as given the average value of voting 'to tell others'). What identifies then the second parameter, σ_ε? The lying moments play an important role in determining the relative variance of signal value σ_{SI}^2 versus the variance of the other reasons to vote σ_ε^2. With small σ_ε, selection into voting is primarily on the signal value, and therefore lying about voting by non-voters should be very limited. If all of the variance is in social-image, the only individuals who will become non-voters are the ones that actually prefer to say that they did not vote (i.e.,, have $s_V < s_N$) and thus would never lie about not voting. With large σ_ε, there is more selection on ε and therefore we expect to see more lying among voters and non-voters.

Given that lying about voting by non-voters is an established fact, this implies that σ_ε^2 cannot be too small. That being said, the estimates will have a hard time separating large values of σ_ε from very large values of σ_ε. Thus, an important robustness check is one in which we fix σ_ε (Table 5, Columns 2 and 3). As we see, the lying cost is tightly linked to the value of σ_ε: with higher σ_ε, the model requires a higher lying cost to match the lying 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. 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. 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. 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. The responsiveness 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.[25] Identifying them separately is not obvious, since they often appear in the model in the product ηS.

Benchmark Estimates. In Table 3, we present the estimates under two assumptions. In the first column, our benchmark specification, we allow for different auxiliary parameters for voters and non-voters, attaining a Weighted Sum of Squared Errors (SSE) of 160.3. In the second column, we force voters and non-voters to have the same auxiliary parameters, leading to a substantially worse fit, with an SSE of 358.8.

Under either set of estimates, the average individual does not take *pride* in saying that they voted: $\mu_V = -3.9$ (se 1.5) in the benchmark and $\mu_V = -6.3$ (se 2.1) in the alternative specification. This reflects the fact that even voters do not sort in to answer the election survey, compared to the non-election survey. We find strong evidence of *stigma* from not voting on average: $\mu_N = -11.3$ (se 1.8) in the benchmark and $\mu_N = -21.7$ (se 3.2) in the restrictive model. This reflects the fact that non-voters strongly avoid the election survey.

The benchmark estimates imply that individuals place an average value of over \$7 ($\mu_V - \mu_N = 7.4$) on being seen as a voter, as opposed to as a non-voter, each time they are asked. There is substantial heterogeneity in these social-image values ($\sigma_{SI} = 9.5$ in the benchmark), implying that 34 percent of individuals would in fact take pride in saying they voted.

Turning to the second key model component, we estimate a cost of lying about voter status of \$7.6 (se 1.2) in the benchmark and \$16.4 (se 2.8) in the more restrictive model. To put this in perspective, we estimate a lying cost of \$7 in a representative cheap talk laboratory experiment (Erat and Gneezy, 2012, see Appendix C). In both settings, a person is lying to a stranger; this cost is likely to be larger in the context of repeated interactions. This significant lying cost, together with the sizable social-image utility, implies that being asked about voting becomes a reason to turnout. We return to this point below.

The distribution of the other reasons to vote ε is estimated to have a large standard devi-

[25] 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).

ation (albeit imprecisely estimated): $\sigma_\varepsilon = 318.7$ (se 691.4) in the benchmark model.

The next panel shows the average value of the key variables—ε, s_V, and s_N—among the individuals who *select* to become voters and non-voters. (Since L is deterministic, it does not differ for the two groups.) The voters differ most from non-voters in the other reasons to vote, ε, with a mean of \$269 for voters and -\$242 for non-voters. Given the high σ_ε, most of the selection into becoming a voter is driven by factors other than social image.

Voters have a more positive social-image value of voting s_V and a more negative social-image value of non-voting, s_N, compared to non-voters. Thus, as expected, voters have larger social-image utility $s_V - s_N$, though the difference is relatively small: \$8.1 for voters versus \$6.2 for non-voters. Given these values, on average non-voters are close to indifferent between admitting they did not vote and claiming they voted ($E^{nv}[s_N] = -10.7$ versus $E^{nv}[s_V] - L = -12.1$), as implied by a lying rate among non-voters of about 50 percent in the control group.

Turning to the auxiliary parameters (bottom panel), we estimate that on average voters display a higher willingness to complete the survey compared to non-voters. (Voters are likely public good providers generally). The estimated time value is \$43 per hour for voters and \$24 for non-voters, a difference consistent with the strong positive correlation between income and turnout (Leighley and Nagler, 1992).[26] Voters and non-voters incur similar social pressure costs from declining to participate in the survey and have similar elasticities of home presence.

Value of Voting to Tell Others. Using the estimates, we compute the value of voting *to tell others*, averaged over the population: $N \int \Phi(s_V - s_N, L)\, dF(s_V, s_N)$. As Figure 1 shows, for positive values of social-image utility ($s_V - s_N > 0$), the value of voting *to tell others* Φ equals $\min(s_V - s_N, L)$, since non-voters can get the social-image utility $s_V - s_N$ by lying (and paying cost L). Thus, the value increases in both the social-image utility and the lying cost.

A key parameter for the value of voting *to tell others* is the expected number of times asked, N, given the assumption that the social image utility cumulates with each additional ask. 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: 61 percent of respondents report being asked at least once, and 15 percent report being asked more than 10 times. On average, respondents report being asked around 5.4 times for the 2010 congressional election, with similar magnitudes for voters and non-voters ($N^v = 5.12$ and $N^{nv} = 6.01$).[27] 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 38 percent of people

[26]The sizeable estimated value of time is consistent with the relatively high median household income of \$92,000 on average across the towns visited.

[27]Respondents were asked to report the number of times asked about the 2010 election by friends, relatives, coworkers, and other people. Total number of times asked is the sum across these categories, each capped at 20 times asked. Similarly, number of times asked about the 2008 election is the number of times asked by friends and relatives, each capped at 20 times asked.

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. (2016).

The top row in Table 4 shows the implied value of voting for voters and non-voters for the 2010 congressional election. The magnitudes are sizable: in the benchmark model, we estimate a total value of voting to tell others of \$18 (se 4.6) for voters and \$13 (se 3.3) for non-voters, with larger estimates in the alternative specification. These magnitudes follow directly from the sizable estimates for the social-image utility and for the lying cost, as well as the fact that people expect to be asked on average five times.

We should point out three important caveats to this measure of the value of voting. First, as in any structural paper, the value depends on the parametric and distributional assumptions; below, we explore how this value varies for a wide set of sensitivity checks. Second, we estimate the social-image utility when asked by a stranger. If social-image concerns or lying costs when interacting with friends, family and colleagues are higher, our estimates are likely to be lower bounds of the social-image value of voting. Third, the estimates assume that the value of being asked is linear in the times asked. If utility is instead concave in times asked, we could be underestimating the value, since we conduct our experiment well after the election, when respondents report already having been asked a number of times.

Even taking these caveats into account, our ability to assign a dollar value to voting through the design is a unique contribution to the literature. For example, 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 for the value of voting are up to a scaling for the voting cost, which is not identified; thus, they do not provide a monetary value of voting due to their model. The unique element in our design that makes the difference is that we use monetary inducements—variation in the value of the survey and incentive to lie about voting—to translate the findings into a monetary value of voting.

A second metric to evaluate the model of voting 'to tell others' is in terms of the extra turnout that it generates, relative to the baseline turnout, 60 percent in our setting. How would turnout change if people stopped asking others whether they voted (that is, $N = 0$)? What if conversely the rate of asking doubled (that is, $2N$ asks)? We can address these questions because we are able to estimate the distribution of other reasons to vote ε.

Table 4 and Figure 9 display the results. With the benchmark estimates, a shift in norms that were to make it undesirable to ask others about voting would lower turnout by 1.9 percentage points. A doubling in the number of asks, perhaps encouraged by political campaigns, would increase turnout by 1.8 points.[28] While these impacts may seem small, consider the extraordinary effort that campaigns put into get-out-the-vote (GOTV) efforts, with the average

[28]The number of times asked can also differ by type of election: in fact, our survey evidence indicates that in the 2008 presidential election people were asked twice as much. Our calculation would imply that just the additional asking (holding the other parameters constant) could contribute 1.8 to 2.5 percentage points.

such letter yielding a turnout impact of 0.2 points (Green, McGrath, and Aronow, 2013).

Get-Out-The-Vote and Welfare. The estimates allow us also to calculate the implied turnout and welfare effects of a GOTV intervention based on informing potential voters that they will be asked whether they voted. This is a GOTV intervention which we designed ourselves and which we evaluate in Section 6. In addition, a related GOTV message, independently designed, was used in the 2012 presidential election and is evaluated in Rogers, Ternovsky, and Yoeli (2016). We assume that this intervention increases the number of times asked to $N+1$.[29]

First, we calculate the welfare effect of this GOTV intervention. For voters, the average value of being asked about voting, $z^v = \max(s_V^v, s_N^v - L^v)$, is estimated to be -2.8 (se 1.2) in the benchmark model. Non-voters are estimated to have a more negative utility from being asked, -5.9 (se 1.5). These sizable negative welfare effects occur because on average individuals do not derive pride from voting, and they strongly dislike admitting that they did not vote.

Second, we predict the effectiveness of the intervention. Given the distribution of the other reasons to vote ε, this GOTV intervention is predicted to increase turnout by 0.3 percentage points (Table 4 and Figure 9). Thus, to get one extra vote with this intervention, 295 people would have to be contacted.

Taking these two numbers together, we estimate that this GOTV intervention would result in a utility cost of \$1189 *per additional vote yielded*. This estimated magnitude is similar in the restrictive model, and implies a disadvantageous cost-benefit analysis, once one takes into account the voter welfare. This welfare disutility dwarfs the cost estimates in the current literature, which typically include just the postal costs of sending the GOTV material (e.g., Rogers, Ternovsky, and Yoeli, 2016). As far as we know, ours is the first welfare evaluation of a get-out-the-vote intervention, an area of vast growth in the political science literature.

Robustness. In Table 5, we explore the robustness of the parameter estimates to alternative assumptions. We maintain the assumption as in the benchmark estimates (reproduced in Column 1) that the auxiliary parameters can differ for voters and non-voters. Online Appendix Table 6 reports the parallel results under the assumption of identical auxiliary parameters.

Before we go into the details, we note some common features. A variety of alternative specifications (with one exception) yield an essentially identical fit of the model (measured as SSE), implying that the alternative models do not affect the ability of the model to explain the data. (This is not true any more if we force the auxiliary parameters to be the same for voters and non-voters, see Online Appendix Table 6.) Also, while most parameter estimates are quite stable, the standard deviation of the other reasons to vote, σ_ε, varies substantially. This is not surprising given that this is the least precisely estimated parameter. This does not much impact the dollar value of voting 'to tell others', but it affects the implied percentage point impact on turnout, which is therefore more variable across estimates.

[29]The impact could be smaller if individuals expect to not be reachable at this follow-up time, or if they miss the flyer (or mailer).

First, we remove the assumption, maintained so far, that lying costs L are deterministic and identical for voters and non-voters. We allow lying costs to have an exponential distribution; this one-parameter distribution ensures that lying costs are always positive. In this specification (Column 2), individuals select into voters and non-voters partly based on the realized lying cost. We estimate the same mean lying cost of \$8 as in the benchmark, with little selection into voting by lying cost (mean lying cost of \$8.4 for voters and \$7.3 for non-voters). We estimate a lower standard deviation of ε than in the benchmark model; as a consequence, there is more selection into voting based on signaling value and a larger implied effect on turnout if potential voters expected to be asked twice as often.

Next, we consider an alternative explanation of the results (Column 3): 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 allow for a utility of talking about politics which is independent of whether one voted or not. With this extra parameter, we lose the ability to estimate a social-image parameter, so we fix μ_V to zero. The estimates of the key parameters of the value of voting are largely unaffected, reflecting the key identifying role of the lying moments, which are unaffected by the introduction of a utility of talking about politics.

We then consider measurement error in the voter records. 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 4, we assume that 10% of respondents in voting households are actually non-voters. Allowing for measurement error leads to similar results as in the benchmark case, with larger turnout impacts of doubling the number of asks. The results are similar if we assume a higher measurement rate of 20% (Column 2 of Online Appendix Table 5) or we allow for symmetric measurement error, and assume that 10% of respondents in a voting (or respectively, non-voting) household are non-voters (respectively, voters) (Column 3 of Online Appendix Table 5).

We also test the robustness of the estimates to the assumed number of times potential voters expect to be asked about voting (Column 5). In case the survey results overstate the number of times asked, we re-estimate the model assuming half as many asks ($N = 2.7$). The model estimates are similar, with, not surprisingly, a lower monetary value of voting 'to tell others' (since people are asked less). We find similar results if we assume voters expect to be asked twice as often (Online Appendix Table 5, Column 4).

Finally, we explore the role of an important parameter that is not precisely estimated, the standard deviation of the other reasons to vote, σ_ε. To demonstrate its role, we fix it to a low level ($\sigma_\varepsilon = 10$, Column 6) and to a high level ($\sigma_\varepsilon = 1000$, Column 7). For low σ_ε, the model has a harder time matching the moments and the model SSE increases from 160 to 172. With low σ_ε, the selection is mostly on the social-image terms, making it harder to match the lying rate. For high σ_ε, instead, the estimates are similar to the benchmark, with a lower implied turnout impact of a doubling in the asks, since fewer people are on the margin of the turnout decision. This highlights the one-sided nature of identification of σ_ε: the data rejects small values of σ_ε, but has a hard time telling apart between sizable to large values of σ_ε.

We consider a number of other robustness checks in Online Appendix Table 5, some of which we discussed above. We obtain similar estimate if we do not drop the observation with no-solicitor sign or other issues (Column 5) or if we include as additional moment the turnout of the treated population in the get-out-the-vote experiment (Column 6).[30] Estimates excluding the moments split by whether households are informed at the door about the election topic (Column 7) indicate a larger social-image utility, not surprisingly, since the excluded moments reveal no differential impact of the social-image manipulation. Estimating the model without the lying incentive moments (Column 8), the heterogeneity in signal parameters is not identified, so we fix $\sigma_{SI} = 10$, as in the benchmark estimates. The estimated signaling parameters μ_V and μ_N are similar to the benchmark, but the lying cost is much less precisely estimated, revealing the key role of the lying incentive treatments.

Estimation with Exogenous Voter Status. In Online Appendix Table 7, we present the results of an alternative estimation method used in previous versions of the paper (DellaVigna, List, Malmendier, and Rao, 2014), which does not model the decision to become a voter or non-voter and instead allows for separate parameters for voters and non-voters ($\mu_V^i, \mu_N^i, \sigma_{SI}^i$ for $i \in \{v, n\}$). Without modeling the turnout decision, we are not able to estimate the distribution of other reasons to vote, ε, and we cannot separately identify the lying cost L. Instead, we estimate of the mean value of saying one voted ($\mu_v = -6$ for voters and $\mu_v - L = -8$ for non-voters) and saying that one did not vote ($\mu_n - L = -28$ for voters and $\mu_n = -7$ for non-voters). The value of voting to tell others is only defined as a function of a given (assumed) value of lying cost. For a lying cost of \$5, this value is \$12 for voters and \$17 for non-voters.

6 Get-out-the-vote Experiment

The experiments described 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 did 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*

[30] To calculate the simulated moment, we assume that potential voters in the GOTV treatment group expect to be asked on average h_0^i times more than they would absent the treatment, where $i \in \{V, NV\}$ is determined by voting behavior absent the GOTV treatment.

election [...] to conduct a survey on your voter participation'. Figure 10 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 10. 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.

Table 6 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 (11,462 received no flyer, 10,805 received a control flyer, and 9,039 received a treatment flyer). 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 (46,868 received no flyer, 23,501 received a control flyer, and 23,436 received a treatment flyer). 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. As described in Section 5, we can compute the predicted increase in turnout due to an increase in N, the number of times asked about voting, for the estimated parameters. Assuming the treatment flyer increases the expected number of asks by 1, the benchmark model predicts an increase in turnout of 0.3 percentage points (Table 4 and Figure 9). This prediction matches the order of magnitude of the findings: it is smaller than the observed effect for the 2010 election (1.3 percentage points), but larger than the effect for the 2012 Presidential elections (0.1 percentage points), and within the confidence interval of both estimates.[31]

The results are consistent with the contemporaneous and independent results of Rogers, Ternovsky, and Yoeli (2016) who similarly inform a treatment group that they may be called after the election about their voting behavior. They also find a positive impact on turnout, of similar magnitude (0.2 percentage points). Importantly, it is intuitive that the effect sizes are much smaller than those 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 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 document substantial shame from admitting to not voting, though little evidence of pride from conversely claiming to vote.

The experimental design allows us to estimate the key social image and lying parameters. We estimate a value of voting simply due to being asked of \$18 for voters and \$13 for non-voters, a sizeable magnitude for a congressional election. These constitute among the first estimates of the private value of voting. We also use the estimates to do a welfare evaluation of a get-out-the-vote intervention, and find that the disutility associated with the intervention is likely to dwarf the administrative cost. This underscores the importance of model-based evaluations of GOTV interventions.

A methodological ingredient of this paper is the tight link between a simple model and the experimental design. 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.

[31]The prediction for the 2012 results should be taken with extra caution given that the social-image and other parameters are likely to vary by election.

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. (2009). "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] Belli, Robert F, Michael W Traugott, Margaret Young, and Katherine A. McGonagle. (1999). *Public Opinion Quarterly*, 63(1), 90-108.

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

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

[9] Bursztyn, Leonardo, 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.

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

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

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

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

[14] DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. (2014). "Voting to tell others." NBER Working Paper No. w19832.

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

[16] Duff, Brian, Michael J Hanmer, Won-Ho Park, and Ismail K White. (2007). "Good excuses: Understanding who votes with an improved turnout question," *Public Opinion Quarterly*, 71(1), 67-90.

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

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

[19] Finan, Frederico and Laura Schechter. (2012). "Vote-Buying and Reciprocity", *Econometrica*, 80(2): 863-882.

[20] 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.

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

[22] Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. (2016). "Why People Vote: Estimating the Social Returns to Voting" *British Journal of Political Science*, 46(2): 241-264.

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

[24] Green, Donald P., Mary C. McGrath , Peter M. Aronow. (2013). "Field Experiments and the Study of Voter Turnout" *Journal of Elections, Public Opinion and Parties* Vol. 23(1), 27-48.

[25] 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.

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

[27] 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.

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

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

[30] Leighley, Jan E. and Jonathan Nagler. (1992). "Socioeconomic Class Bias in Turnout, 1964–1988: The Voters Remain the Same, " *American Political Science Review*, 86: 725–736.

[31] Levine, David K., and Thomas R. Palfrey. (2007). "The paradox of voter participation? A laboratory study. " *American Political Science Review* , 101(01), 143-158.

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

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

[34] Passarelli, Francesco and Guido Tabellini. Forthcoming. "Emotions and Political Unrest" *Journal of Political Economy.*

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

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

[37] Rogers, Todd, John Ternovski, and Erez Yoeli. (2016). "Potential follow-up increases private contributions to public good." *PNAS*.

[38] 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.

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

A Appendix A - Mathematical Appendix – For Online Publication

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

Turning to the probability of answering a survey, conditional on opening the door, an individual will agree to the survey if $s+m-c+z \geq -S$ assuming she knows that the survey has an election topic and if $s+m-c \geq -S$ in case she does not know. By the same token as above, holding constant the selection into opening the door, 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 $P\left(SV\right)^v_I \geq P\left(SV\right)^v_{NI}$ under *Pride* and $P\left(SV\right)^{nv}_I \leq P\left(SV\right)^{nv}_{NI}$ 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 opening the door. We consider separately the following four exhaustive cases: (i) $max(s+m-c+z, s+m-c) < -S$. In this case, $P\left(SV\right) = 0$ under any condition; (ii) $\min(s+m-c+z, s+m-c) \geq -S$. In this case, the person will complete the survey conditional on opening the door, so $P\left(H\right) = P\left(SV\right)$, and the comparison follows from the results above on $P\left(H\right)$; (iii) $s+m-c+z < -S \leq s+m-c$. 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) $s+m-c < -S \leq s+m-c+z$. 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 $P\left(SV\right)^v_{FE} \geq P\left(SV\right)^v_F$. Under *Stigma*, cases (i), (ii), and (iii) apply and pairwise comparisons for all these cases show $P\left(SV\right)^{nv}_{FE} \leq P\left(SV\right)^{nv}_F$.

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 $s_N - L + I \geq s_V$ or $-\left(s_V - s_N\right) - L \geq -I$. Under the assumption $s_V - s_N > 0$ 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 $s_V - L \geq s_N + I$ or $\left(s_V - s_N\right) - L \geq I$. The left-hand side can be positive or negative depending on whether the net signaling 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 $-\left(pV + g - c\right)$, we can rewrite the probability of voting as

$H\left[N\left[\max\left(s_V, s_N - L\right) - \max\left(s_N, s_V - L\right)\right]\right]$. Under the assumptions $s_V - s_N > 0$ and $L > 0$, it follows that $\max\left(s_V, s_N - L\right) = s_V$ and that $s_V > \max\left(s_N, s_V - L\right)$. Hence, the term in square brackets is positive and the conclusion follows.

B Appendix B - Estimation Appendix

The simulated method of moments estimator chooses the parameters $\hat{\xi}$ that minimize the distance given by $\left(m_N\left(\xi\right) - \hat{m}\right)' W \left(m_N\left(\xi\right) - \hat{m}\right)$, where $m_N\left(\xi\right)$ are the simulated moments given parameters ξ for N potential voters and $\hat{m}$ are the estimated empirical moments. In our benchmark estimations with auxiliary parameters that differ across voters and non-voters, we calculate the simulated moments with $N = 750,000$ potential voters. For benchmark estimation with auxiliary parameters that are the same across voters and non-voters, we use at least $N = 500,000$ potential voters. 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.)

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, location-day, and hour-of-day. The lying moments are conditional on location-day fixed-effects, given the smaller sample of survey respondents. 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. We assume zero covariance between the following sets of moments: door opening, survey completion by treatment, and opting out; survey completion by whether respondent was informed about survey content; lying; and turnout.

The simulated method of moments estimator using weighting matrix W achieves asymptotic normality, with estimated variance

$$(\hat{G}'W\hat{G})^{-1}(\hat{G}'W(1 + J_m/J_s)\hat{\Lambda}W\hat{G})(\hat{G}'W\hat{G})^{-1}/N,$$

where $\hat{G} \equiv N^{-1}\sum_{i=1}^{N}\nabla_{\xi}m_i(\hat{\xi})$, $\hat{\Lambda} \equiv Var[m(\hat{\xi})$, J_m is the number of empirical observations used to calculate a moment, and J_s is the corresponding number of simulated observations used for the moment (Laibson, Repetto, and Tobacman, 2007). 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 constrained nonlinear minimization routine implemented in Matlab as the fminsearchbnd routine. We impose the following constraints: $\mu_j \in [-100, 100]$ for $j \in \{s_V, s_N\}$ (finite social-image utilities), $\sigma_{SI} \in [0, 100]$ (positive standard deviation of social-image utilities), $L \in [0, 50]$ (non-negative lying costs), $S_s \in [0, 100]$ (social pressure non-negative), $\mu_s \in [-100, 100]$ (finite value of doing a survey), $\sigma_s \in [0, 100]$ (positive standard deviation of value of doing a survey), $h_0, r \in [0, 1]$ (probabilities between zero and one), $\eta \in [0, 0.5]$ (finite responsiveness of opening the door), $v_s \in [0, 200]$ (finite and non-negative value of time), $\mu_\varepsilon \in [-500, 500]$ (finite mean non-signaling value of voting), and $\sigma_\varepsilon \in [0, 500]$ (positive standard deviation of non-signaling value of voting).

Only two of these constraints appear to impact the estimation. First, the model cannot distinguish between large values of η, the responsiveness of opening the door. For $\eta > 0.5$, the cost to change the probability of opening the door is negligible, and therefore everyone chooses to be home or away with certainty. Second, as discussed in Section 5, the identification of

σ_ϵ is one-sided: there is little difference in the simulated moments for large values of σ_ϵ. By restricting the search space for η and σ_ϵ, we aid the optimization routines without qualitatively changing the results.

We begin each run of the optimization routine by quasi-randomly choosing a starting point. First, candidate start points are randomly drawn from a uniform distribution over a more targeted parameter space: $\mu_{S_V} \in [-20, 20]$, $\mu_{S_N} \in [-30, 10]$, $\sigma_{SI} \in [0, 30]$, $L \in [0, 20]$, $S_s \in [0, 10]$, $\mu_s \in [-50, 0]$, $\sigma_s \in [0, 50]$, $h_0, r \in [0.2, 0.4]$, $\eta \in [0, 0.5]$, $v_s \in [0, 100]$, $\mu_\varepsilon \in [-30, 100]$, and $\sigma_\varepsilon \in [50, 200]$. To aid the optimization, we restrict the set of randomly selected starting points to those with parameter values that imply turnout of 40-80%. To avoid selecting local minima, we choose the run with the lowest minimum squared distance of 720 runs in the model with auxiliary parameters that vary by voters and non-voters (and at least 480 runs in the model with auxiliary parameters that are the same).

C Appendix C - 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 8, 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 8 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.

D Appendix D - 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.

We attach the entire content of the survey for the 4 condition: (i) 10-minute survey with no incentive to lie; (ii) 10-minute survey with 8-minute incentive to lei; (iii) 5-miute sruvey with no incentive to lie; (i) 5-minute survey with $5 incentive to lie.