Experimental design and pre-analysis plan for "Giving social pressure: Costs, motives, and effects"

Woojin Kim*

September 7, 2020

1 Research questions

1.1 Motivation and overview

Suppose you find out that one of your peers hasn't registered to vote for an upcoming election—do you tell them they should? Anecdotally, many people would be hesitant to do so for many possible reasons. They may think it's none of their business, they may believe that saying something still won't change the other person's behavior, or they may not want their peer to feel pressured. On the other hand, most of us can perhaps think of someone who is likely to speak up in that situation.

This project captures the psychological cost (or value) of giving social pressure in an experiment featuring the upcoming 2020 U.S. General Election. It focuses on voter registration, which is fundamental to American elections yet understudied compared to turnout. In the experiment, registered college students face the preceding situation, where they are asked to email an unregistered peer. The experiment elicits their incentivized willingness to pay (WTP) to (not) message and pressure their peer to register.

The study not only quantifies the cost or value to send a message, it also decomposes the WTP into the Self- and Other-regarding sources. Self-regarding motives are numerous, and

*UC Berkeley, Dept. of Economics, woojin@berkeley.edu

are defined as those that involve personal gain or loss. For instance, participants may not want to pressure others because composing a message takes effort, or there is a chance of confrontation or rejection; on the other hand, they may want to pressure others to express their disapproval, or to signal that they value the norm.

Meanwhile, Other-regarding motives consider the welfare of others. These include altruism (as the recipient may dislike being pressured) and civic duty (to increase electoral participation for the good of society). This project includes a survey instrument to directly elicit incentivized beliefs relevant to these Other-regarding motives. The survey asks potential senders how recipients would respond to their message, in particular, whether the recipients would like the message and have a higher chance of registering to vote. To capture the Other-regarding motives, the study analyzes how much of the WTP can be explained by these beliefs.

As the experiment randomizes whether an unregistered participant receives a message, it can uncover the causal impact of this peer-to-peer social pressure communication on registering to vote. Furthermore, the rich set of variables collected in the surveys can offer insights into interesting dimensions of heterogeneity. For instance, is it easier to pressure a friend or a stranger? For whom is social pressure more effective? How does the willingness to pressure vary by the sender's gender, or the recipient's gender?

In addition to these reduced-form findings, I build a theoretical framework to identify the channel through which these social pressure messages work (if at all). Under an action-based social signaling model, there are three mechanisms: beliefs about the norm, sensitivity to the norm, and costs/benefits of taking the action. The model can discern which of these change after receiving the message.

Continuing under this framework, I then forecast the outcomes of a policy that scales this intervention of asking registered students to message their unregistered peers to a campus-wide level. This structure allows us to calculate the total welfare cost of social pressure from both the recipient and sender's sides, and the cost-effectiveness of this policy as beliefs converge in equilibrium.

1.2 Primary research questions

1. What is the cost or value of giving social pressure?

(a) What are the sources of this cost or value? How does it decompose into Self- and Other-regarding sources?

(b) How does this cost or value depend on the relationship between the sender and receiver (e.g., friends or strangers)?

(c) How does this cost or value vary by the gender and race/ethnicity on both the sender and receiver's sides?

2. How does receiving social pressure affect the compliance to the norm?

(a) What are the mechanisms behind this effect (if any)?

(b) How does this effect depend on the relationship between the sender and receiver (e.g., friends or strangers)?

(c) How does this effect vary by the gender and race/ethnicity on both the sender and receiver's sides?

(d) What would be the equilibrium outcomes of a social pressure campaign for voter registration?

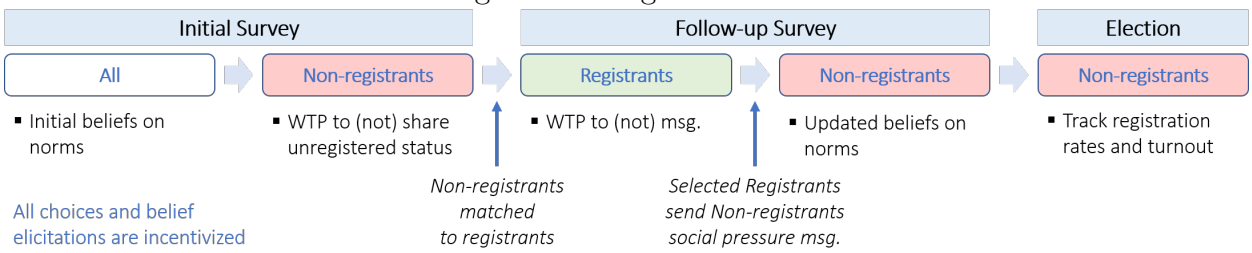
1.3 Secondary research questions

1. Does pressuring others increase one's own chances of complying with the norm?
2. Do people have correct beliefs about other people's compliance with the norm?
3. Do people know how effective giving social pressure is and how it will be received?
4. How do people compose social pressure messages?

2 Experimental design

Logistics and overview This study will be conducted online from September 8 to October 19, which is the deadline in California to register to vote for the 2020 U.S. General Election. Students attending UC Berkeley, UC Santa Barbara, and UCLA will be invited to participate. I will recruit student participants from existing subject pools at behavioral science labs (Xlab at UC Berkeley, EBEL at UCSB, and the Anderson Behavioral Lab at UCLA). If more participants are needed to reach a sample size of 1000, I will also advertise

Figure 1: Design overview



the study through campus email lists. The study will be conducted in several rounds, which have been scheduled as follows:

- Round 1 (Pilot): Launch September 8, 2020 at UC Berkeley (Xlab)
  - Fall semester instruction begins at UC Berkeley on August 26, 2020.
  - Capped at ~150 participants.
- Round 2: Launch September 21, 2020 at UC Berkeley
  - If the 150 participant quota is not reached through Xlab in Round 1, then I will also ask instructors of large courses (e.g., Econ 1) to advertise the study to their enrolled students.
  - Capped at 650 additional participants.
- Round 3: Launch October 5, 2020 at UC Santa Barbara (EBEL) and UCLA (Anderson Behavioral Lab, approval pending)
  - Fall quarter instruction begins at UCSB and UCLA on October 1, 2020.
  - I will accept participants as long as funds remain.

In each round, participants will take two Qualtrics surveys within 10 days: (1) the Initial Survey and (2) the Follow-up Survey. There are two types of participants. "Registrants" are those who have already registered to vote when they begin the study, and "Non-registrants" are those who have not. Figure 1 shows an overview of the design.

Initial Survey (open for 3 days) The Initial Survey first screens for eligibility to vote and to register online in California. Participants then give their Informed Consent. The Consent Form states that the purpose of this project is broadly "to study peer communication among

students about the election." For this purpose, the Consent Form mentions that participants may share their full name, voter registration status, and email address with other student participants in the study, who may send them an email about the election.

Next, the survey collects student and demographic information. The survey then asks participants whether they were registered to vote *before* the Initial Survey was launched. This is to prevent participants from registering to vote after they start the survey and then reporting that they have registered. All participants report their incentivized beliefs on their campus's voter registration rate, which measures their perception of the norm. They are also asked what they think are the chances the Democratic and Republican parties gain a majority in the Senate and House respectively following the election. This second question that asks for their beliefs on the outcome of the election is to draw attention away from voter registration. As much as possible, I intend for "untreated" participants (i.e., those who do not receive a social pressure email message about registering to vote and those who are not asked to send one) to remain unaware that voter registration is the primary outcome of the study, which would help to mitigate experimenter demand effects.

The survey branches for Non-registrants, who have not registered to vote yet. They provide their willingness to pay (WTP) to (not) share their registration status with other participants.[1] They are informed that if their status is shared, they may also receive an email from another participant about the election. To reduce experimenter demand effects, the topic of the email is merely described as being "about the election", and voter registration is not specified. The WTP is elicited through a standard multiple price list mechanism. The WTP is incentivized from a range of -$7 to $7 in increments of $1. The price list is presented in the following format:

	Choose left	Choose right	
Row #) Share my info and get paid $\$x$	∘	∘	Don't share my info and get paid $\$y$

Rows 1-8 have x constant at $5 and y decreasing from $12 to $5 in increments of $1. From Rows 8-15, y remains constant at $5 and x increases from $5 to $12 in increments of $1. On this 15-row price list, a Non-registrant can indicate either a positive or negative WTP to share their name and registration status. The WTP is defined as the difference in compensation between the left and right options at the row where the Non-registrant switches from the

[1] Their WTP responses capture the welfare cost of revealing to others that they have not registered to vote, which could constitute a loss in social recognition. This welfare cost features in the structural estimation.

right side to the left side (since the right option is decreasing in compensation down the table, whereas the left option is increasing). To simplify the decision and ensure that choices are consistent, participants fill out the rows through a step-by-step process of elimination, which takes 4 steps. For example, first participants indicate their preference for Row 8 where $x = y = \$5$. Their answer implies a consistent choice for 7 other rows (e.g., if "Choose left" is chosen in Row 8 when $x = y = \$5$, then"Choose left" should also be selected for Rows 9-15 where $x > \$5, y = \$5$). This process continues until the row where the participant would like to switch from the right option to the left option (if ever) is identified. If participants choose all left or all right options, this implies that their WTP to share their information is either less than -$7 or greater than $7. In these censored cases, I ask for their hypothetical WTP beyond these bounds.

The choices on the price list are incentivized under the standard mechanism, where either (A) one of the rows will be randomly selected and the participant will get their preferred option in that row, or (B) the computer will not consider their choices on the price list, instead randomly assign them to a condition (share or not share info), and compensate the participant the maximum amount ($12). They are told that this assignment to (A) or (B) will happen after they submit their Initial Survey. After the WTP exercise, they are invited to refer their campus friends to participate in the study before finishing the survey. Registrants (who have already registered to vote) skip the WTP elicitation and go straight to the referral prompt on the last page of the survey.

After the Initial Survey closes, I download the data from the Qualtrics server. I spend one day verifying the information reported by participants and performing the randomization described in Section 3. For UC Berkeley and UCLA student participants, I check whether a student with the same full name or email address exists in the online public UC Berkeley directory or UCLA directory. For UC Santa Barbara student participants, I search for a matching student with the same full name and date of birth on the online public UCSB student verification portal. For all student participants, I also check the universe of California state voter registration records for someone with the same full name and date of birth. I ensure that those who state they have registered can be found in the records, and those who state they have not registered cannot be found in the records. Since I have the California voter file updated as of August 7, 2020, the survey also asks those who state that they have registered whether they did so in the last two months. If they state that they registered in the last two months but I cannot find them in the voter file, I give them the benefit of the doubt. I can verify whether they were in fact registered to vote when they took the Initial Survey using the updated voter file that I will request following the election. (The voter file

records the registration date.)

Follow-up Survey: Registrants (open for 3 days) The Follow-up Survey is first sent to Registrants, who are shown the names and registration statuses of three other participants. They are first asked to characterize their relationship with each of the participants as Strangers, Acquaintances, Friends, or Close friends. The Registrants are told they may be randomly selected to email one of the three potential recipients. The purpose of the email randomly differs between Registrants and can be one of two types. The "treatment" email pressures the recipient to register to vote ("Pressure Message"). The "control" email informs the recipient about legislative districts in California ("Info Message").[2] Registrants are shown a template of whichever message they were assigned, and are informed that if they are selected to send the message, they can edit it within general guidelines.

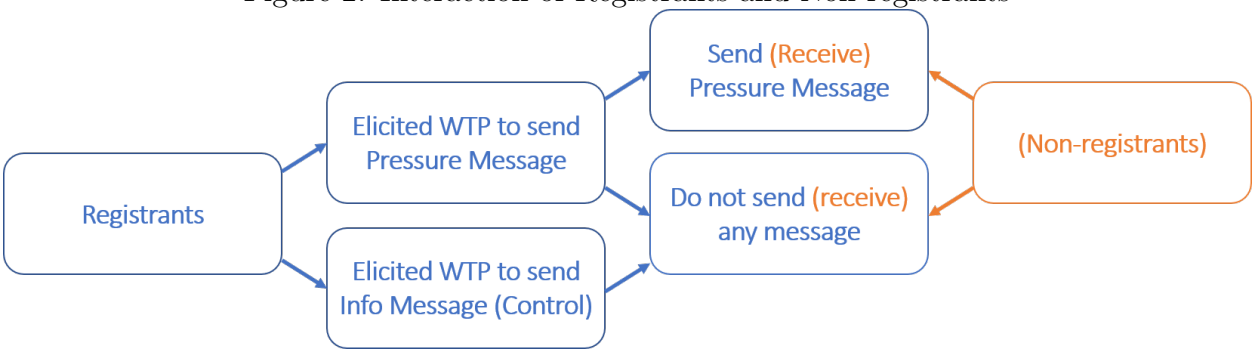
Registrants are also told that if they are selected to send the message, they will have to send it either directly or anonymously. A "Direct Message" must be sent by the Registrant from their own email account, while an "Anonymous Message" is sent by the research team without mentioning the Registrant's name. For each of the three potential recipients, Registrants are shown whether they would have to send a Direct or Anonymous Message if they are selected to message that recipient. To summarize, the Pressure vs. Info Message condition varies *between* Registrants, while the Direct vs. Anonymous Message condition varies *within* Registrants.

For each of the three potential recipients, Registrants are asked on a 7-point agree/disagree Likert scale whether they think the recipient would like to receive the message from them. Then, they are asked the chances that the recipient registers to vote by the election (or for the Info Message condition, the chances that the recipient knows who their local representatives are) first if they do not send the message, and then if they do send the message. For Registrants assigned to the Pressure Message condition, all these beliefs elicitations are incentivized.

Next, the Registrants state their WTP to (not) send the email message via the same multiple price list mechanism as for the Non-registrants on the Initial Survey. Therefore, they fill out 15 rows $\times$ 3 participants = 45 rows in total (by answering 4 process-of-elimination questions

[2]Registrants assigned the Pressure Message will only see Non-registrants as potential recipients. If the proportion of Registrants in the sample is greater than 75%, then Registrants assigned the Info Message may see a mix of Registrants and Non-registrants as potential recipients. Otherwise, they will only see Non-registrants as potential recipients as well.

Figure 2: Interaction of Registrants and Non-registrants



per potential recipient).

For the Registrants who are selected to send a message, they must compose their message on the survey. For Direct Message senders, they must also email their message directly from their campus email account (BCC'ing me) before finishing the study. For all other Registrants, they must send me a confirmation email before submitting their responses. This confirmation email serves to measure the attrition at this stage simply from the hassle of opening the email client and sending *any* email, not just a message to another participant. They are informed that if they do not follow all of these steps, their participation will be deemed incomplete, and they will not be offered any compensation. Figure 2 shows the assignment and interaction of Registrants and Non-registrants.

Follow-up Survey: Non-registrants (open for 3 days) After the Registrants have taken the Follow-up Survey, the link is sent to Non-registrants. Non-registrants are asked whether they received an email about the election from another student participant. If they did, they are asked to describe the contents of the email, and then how much they liked/disliked receiving it on a 7-point agree/disagree Likert scale. If they were assigned to receive a Direct Message, they are asked whether the sender was a Stranger, Acquaintance, Friend, or Close Friend, or if they do not remember who send it. Lastly, the Non-registrants may update their guesses on their campus's registration rate and outcomes of the election.

Post-election Non-registrants' registration statuses and turnout in the election are tracked to assess the causal impact of receiving a social pressure message.

3 Randomization

Randomization in this study is an involved process, since there are interventions on both Non-registrant and Registrant sides, and the two groups also interact. The procedure is conducted in the following steps after data from the Initial Survey have been collected.

For 10 percent of Non-registrants, their WTP choices on the Initial Survey to share their name and voter registration status are taken into consideration; a random row on the multiple price list will be selected, and they are assigned their preferred option. (These Non-registrants with endogenous sharing/not sharing are excluded in the analysis for the effect of pressure messages on voter registration and turnout.) For the other 90 percent, they share their full name and registration status with Registrants on the Follow-up Survey by default.

One research question is the differential effect of being pressured by a friend as opposed to a stranger. From the Initial Survey, I can only tell if participants are friends if they referred one another. However, I expect the number of referrals to be limited. Hence, if the proportion of Non-registrants who referred or were referred by a registered participant is less than 50%, I assign a lower chance of optional sharing to those Non-registrants. In particular, I generate a random number between $[0, 1]$ for all Non-registrants. Let $p(< 0.5)$ be the proportion of Non-registrants who are connected to a registered participant by referral. Those Non-registrants whose random number is less than $\frac{p}{5}(< 0.1)$ are assigned to share or not share their info based on their choice in a random row on the price list. Non-registrants without any registered referral connections have a higher chance of optional endogenous info-sharing. They are assigned their choice on a randomly selected row if their random number is less than $\frac{0.1-\frac{p^2}{5}}{1-p}(> 0.1)$.

If p is less than 0.5, Non-registrants who are connected to a registered participant by referral will have a higher chance of being messaged, especially by their registered friend. 5/8 of these Non-registrants whose sharing is not endogenous will be randomly assigned to receive a pressure message from their friend. The remaining 3/8 will be assigned to receive or not receive a message from a random participant in the same process for other non-optional Non-registrants as described below.

The remaining non-optional Non-registrants (who either are not connected to a registered participant by referral or were not selected to be messaged by their registered friend) are assigned to receive or not receive a pressure message by stratified randomization. I partition the sample of Non-registrants into the following bins:

- 3+ bins for WTP to (not) share status depending on the variance in the distribution (e.g., lower half with $WTP < 0$, upper half with $WTP < 0$, and $WTP \geq 0$)
- 2+ bins for gender (e.g., male and female/other)
- 3+ bins for race/ethnicity (e.g., White, Asian, and other)
- 2+ bins for age (e.g., below median and above median)

The example partition above would generate $3 \times 2 \times 3 \times 2 = 36$ strata. Following Duflo, Glennerster, and Kremer (2008), I ensure that each strata has at least 4 Non-registrants. Within each strata, half the Non-registrants are assigned to receive a pressure message, and the other half are not. Then, those Non-registrants assigned to receive a pressure message from a registered non-referred participant are randomly matched to a registered participant who is either (a) same gender and race/ethnicity, (b) same gender and different race/ethnicity, (c) different gender and same race/ethnicity, or (d) different gender and race/ethnicity. The randomly selected Registrant "senders" are thus assigned to the Pressure Message condition, will see their matched Non-registrant recipient under the Direct Message condition, and will have to send the Direct Pressure Message to their matched recipient.

Optional info-sharing Non-registrants are randomly matched to any remaining registered participant who has not been assigned to send a message. These optional Non-registrants are randomly assigned to either the Direct Info, Anonymous Info, Direct Pressure, or Anonymous Pressure condition. Therefore, there is a small chance (which depends on the number of optional Non-registrants that end up sharing their status) that messaging a recipient under any condition is optional. Nevertheless, most Registrants assigned an Info Message or an Anonymous Message will not actually send the message. Registrants do not know about these chances when they take the study.[3]

Next, the remaining Registrants who have not been selected as senders are randomly assigned to the Info Message treatment until 25-33% of all Registrants have been assigned to the Info Message treatment. (The exact proportion depends on the overall proportion of Registrants in the sample to balance the number of times a Non-registrant's information is shown to other participants.) To maximize the power of analyses *within* the Pressure Message treatment, the rest of the unassigned Registrants are designated to the Pressure Message treatment, but will not actually send a message. Then, all Registrants randomly draw Non-registrants until they have each been matched with three Non-registrants in total. (They draw two more

[3]Even if they did, reporting their true preferences for the WTP choices remains incentive compatible.

Non-registrants if they have already been assigned to actually send a message). The drawing procedure is coded so that Non-registrants are shown under both the Direct and Anonymous Pressure Message conditions to two different Registrants, and at least once under the Info Message condition (which could be Direct or Anonymous). The number of times a Non-registrant is drawn is kept as even as possible across Non-registrants. All Registrants see at least one recipient under the Direct Message condition, and at least one recipient under the Anonymous Message condition.

This randomization procedure ensures that a Non-registrant can only receive one message (if any), and that a Registrant can only send one message (if any). For Non-registrants whose info-sharing is not endogenous, if they receive a message, it will be a Direct Pressure Message about registering to vote. That means that most Registrants who actually send a message will send a Direct Pressure Message. The rationale is to maximize the power (which is already low) for detecting an effect of receiving a pressure message on voter registration rates.

4 Pre-analysis plan

This section lays out the empirical analyses. For each of the research questions, I describe the construction of the outcome variable and the covariates, and the empirical specification.

4.1 What is the cost or value of giving social pressure?

4.1.1 Primary outcome variable

The primary outcome variable of interest in this study is the Registrants' WTP to (not) send the messages on the Follow-up Survey. This variable ranges from -$7 to $7 in increments of $1 and is elicited via the multiple price list. It is defined as the compensation on the left option minus the compensation on the right option at the row where the Registrant switches to the left option from the right option. For example, a Registrant who states the following preferences has a WTP to send the message of -$3.

Rows 1-3) Don't send the message and be paid \$5	$\prec$	Send the message and be paid $\$y \geq 10$
Row 4) Don't send the message and be paid \$5	$\prec$	**Send the message and be paid \$9**
Row 5) Don't send any message and be paid \$5	$\succ$	**Send the message and be paid \$8**
Row 6-15) Don't send any message and be paid $\$x \geq 5$	$\succ$	Send the message and be paid $\$y \leq 7$

Censoring Censoring may be an issue since WTP is only incentivized from -\$7 to \$7. For Registrants who state a WTP at the bounds, the survey also asks for their hypothetically unbounded WTP.[4] If more than 10% of the WTP distribution is on either of the bounds, then I run two robustness checks for all specifications in the analyses. First, I use a Tobit model. Second, I use the hypothetically unbounded WTP as the outcome variable for censored responses. To account for outliers, I winsorize within the hypothetical WTP less than -\$7 at the 10th percentile, and within the hypothetical responses greater than \$7 at the 90th percentile.

Attrition For Registrants who fill in the Follow-up Survey until they are selected to send a message but then either close or submit the survey without sending the message, I interpret their WTP to send the message to be censored at whatever amount they would have been compensated for sending the message and completing the study. For example, if a Registrant would have been compensated \$12 for sending the message and completing the study but chose not to send the message, I interpret his or her WTP to be $\leq -\$12$.[5]

Furthermore, I can discern whether the attrition at this stage is to avoid sending *any* email (for hassle costs) as opposed to the social pressure email in particular. Registrants who do not have to (directly) send an email message to another participant still have to send the research team a confirmation email. If the attrition is solely due to the hassle cost of opening the email client and sending an email, then the attrition rates should be similar, regardless of whether the Registrant has to email another participant or the research team.

[4]For example, if Registrants prefer to not send the message and get paid \$5 than to send the message and get paid \$12, the survey asks how much they would have to be compensated to send the message, rather than to not send the message and get paid \$5. Whatever amount they state minus \$5 is their hypothetical WTP.

[5]I use a Tobit model to include these cases of attrition with a lower bound of -\$12. For the participants who completed the study but had their incentivized WTP censored at -\$7, I use their hypothetical WTP if it is between -\$12 and -\$7. If it is less than -\$12, I set it to -\$12 and also include it as censored.

4.1.2 Covariates

- $\Delta\hat{a}_j^i$ is sender i's predicted percentage point effect of the message on the chances that recipient j does the targeted action (either finds out who their local representatives are for the Info Message, or registers to vote for the Pressure Message). The Follow-up Survey asks the potential sender the percent chances that the recipient does the targeted action without a message, and then with a message. $\Delta\hat{a}_j^i$ is the percentage point difference between the two predictions.

- $\hat{u}_j^i$ is sender i's response to the following question on a 7-point Likert scale ranging from "Strongly disagree" to "Strongly agree": *"[Recipient j] would like to receive my [Anonymous/Direct] [Info/Presssure] Message about [local legislative districts/registering to vote]"* . It is transformed to a discrete measure from -3 to 3, where -3 is "Strongly disagree", 3 is "Strongly agree", and 0 is "Neither agree nor disagree".

  - $\hat{u}_j^i$ enters non-parametrically in Specification 4 with indicators for each level of the Likert scale (e.g., "Strongly disagree", "Disagree",...). For reliable estimation, there should be enough (at least 50) responses in each level. Levels with no responses can be dropped, but levels with nonzero yet an insufficient number of responses will be grouped with an adjacent level. For example, if "Strongly disagree" only has 10 responses, "Disagree" has 100, and "Somewhat disagree" also only has 10, I will group all three together under the same level "Disagree".

- $\text{Reln}_{ij}^f, f \in \{S, A, F, CF\}$ is a variable for the relationship between the sender i and recipient j. There are four possible categories: Strangers, Acquaintances, Friends, and Close friends. The number of non-Stranger relationships are expected to be limited. If either the Acquaintances or Close friends category has less than 30 observations, it will be merged with the Friends group.

- $G_{i/j} \in \{M, F/O\}$ is an indicator for male or female/other. For example, $M_i \times F_j = 1$ means a male sender i and a female/other recipient j.

- $R_{i/j} \in \{W, A, O\}$ is an indicator for each race/ethnicity group (e.g., White, Asian, and other). The exact groups will be determined during randomization to partition the sample evenly.

- Diff. race_{ij} is an indicator for whether the sender i and recipient j are of different race/ethnicity.

4.1.3 Empirical specification

What is the cost (or value) of giving social pressure? The baseline specification simply compares the average WTP to send the message across the four intervention groups. To begin, I run the pooled OLS regression:[6]

$$WTP_{ij} = \beta_0 + \beta_1 \text{Direct}_{ij} + \beta_2 \text{Pressure}_i + \beta_3 \text{Direct}_{ij} \times \text{Pressure}_i + \varepsilon_{ij} \quad (1)$$

where i indexes the sender and j the recipient. Direct_{ij} is the indicator for sender i being assigned to potentially send recipient j a Direct Message (as opposed to an Anonymous Message), and Pressure_i is the indicator for sender i being assigned the Pressure Message about registering to vote (as opposed to an Info Message about local legislative districts). In all specifications, I cluster standard errors by sender.

The primary two-sided null hypothesis is $\beta_0 + \beta_1 + \beta_2 + \beta_3 = 0$, i.e., that there is no cost or value to send a Direct Pressure Message.

What are the sources of this cost or value? By decomposing the WTP across the four intervention arms, Specification 1 already begins to probe the sources of this cost or value. There are no prior studies to predict the sign and magnitude of the coefficients, which may be positive or negative for a number of reasons. A few possibilities are listed below.

- The constant β_0 estimates the average WTP to send an Anonymous Info Message. I expect this to be close to \$0.
- $\beta_0 + \beta_1$ estimates the WTP to send a Direct Info Message. β_1 could be negative from a desire for privacy, or may be positive to signal to general interest in politics.
- $\beta_0 + \beta_2$ estimates the WTP to send an Anonymous Pressure Message. β_2 could be negative if senders dislike telling other people what to do, or could be positive if they like to signal to themselves that they are civically minded.

[6] Since the Info vs. Pressure Message condition only varies between senders, I cannot estimate Specifications 1 and 2 under a standard fixed-effects model without dropping β_0 and β_2. To account for individual heterogeneity and still estimate these coefficients, I can run a dummy variable regression with indicators for each sender, and then compare the average of the fixed effects between the Info and Pressure Message groups. The average of the fixed effects in the Info Message group is an estimate of β_0, and the difference is an estimate of β_2.

- $\beta_0 + \beta_1 + \beta_2 + \beta_3$ estimates the WTP to send a Direct Pressure Message. β_3 could be negative if senders believe pressuring may lead to confrontation, or could be positive if they derive utility from signaling to the recipient that they care about electoral participation.

I test each of the coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ against the two-sided null that they are equal to 0 and compare their magnitudes. These tests broadly identify the source of the WTP outcome. For instance, if $\beta_0 = \beta_1 = \beta_2 = 0$ but $\beta_3 < 0$, then pressuring has a cost, but only when senders have to reveal their identity to the recipient. On the other hand, if $\beta_0 = \beta_2 = \beta_3 = 0$ but $\beta_1 < 0$, then directly communicating with another participant—not pressuring per se—is costly.

The next regression checks whether any differences among the four intervention arms can be explained by how effective the sender predicts the message will be ($\Delta\hat{a}^i_j$) and how much the sender thinks the recipient will like receiving the message ($\hat{u}^i_j$). In particular, I run the regression:

$$\begin{aligned} WTP_{ijct} = \beta_0 &+ \beta_1 \text{Direct}_{i\to j} + \beta_2 \text{Pressure}_i + \beta_3 \text{Direct}_{ij} \times \text{Pressure}_i \\ &+ (\theta_1 + \theta_2 \text{Pressure}_i)\Delta\hat{a}^i_j + (\alpha_1 + \alpha_2 \text{Direct}_{ij})\hat{u}^i_j + \omega_c + \tau_t + \varepsilon_{ijct} \end{aligned} \tag{2}$$

where ω_c, τ_t are campus and time (i.e. round of study) fixed effects. Under a standard expected utility framework, the coefficients on $\Delta\hat{a}^i_j$ estimate how much the sender values the recipient taking the action. Since senders may value informing the recipient about their local representatives differently from getting the recipient to register to vote, a separate coefficient is included for the Pressure Message condition.

α_1 represents altruism and estimates the sensitivity of the senders' WTP to how much they think the recipient would like the message. As constructed in Section 4.1.2, $\hat{u}^i_j$ is negative (positive) when the sender thinks the recipient will dislike (like) the message, and is normalized to 0 if neither is the case. α_2 allows the slope on $\hat{u}^i_j$ to differ when the sender is identified to the recipient. Senders who are purely altruistic would have $\alpha_1 > 0$ and $\alpha_2 = 0$—they should not care whether the message is direct or anonymous. Senders who are sensitive to whether the recipient would like/dislike the message only for Self-regarding concerns (e.g., a lower $\hat{u}^i_j$ could mean a higher chance of confrontation) may have $\alpha_1 = 0$ and $\alpha_2 > 0$.

The main purpose of Specification 2 is to check whether $\beta_2 = 0$ and $\beta_3 = 0$. That is, are

there any differences in the WTP to send a Pressure Message as opposed to a non-pressuring Info Message that is not explained by the predicted effects on the recipient?[7]

How does the cost or value of giving social pressure decompose into Self- and Other-regarding sources? The subsequent analysis uses only the responses from potential senders of the Pressure Message, which has the advantages of including sender fixed effects and having all belief variables incentivized.[8] First, I re-run Specification 2 under a fixed-effects regression:[9]

$$\begin{aligned} WTP_{ij} = \beta_i &+ \beta_1 \text{Direct}_{ij} \\ &+ \theta \Delta \hat{a}_j^i + (\alpha_1 + \alpha_2 \text{Direct}_{ij}) \hat{u}_j^i \\ &+ \varepsilon_{ij} \end{aligned} \tag{3}$$

where β_i is the sender fixed effect (which absorbs campus and time fixed effects). Defining $\beta_0 \equiv \bar{\beta_i}$, I can separate the WTP into Self-regarding motives (captured by $\beta_0, \beta_1, \alpha_2$) and Other-regarding sources (captured by θ, α_1).

The linear slopes on $\hat{u}_j^i$ in Specification 3 is a restriction. If the sender fixed effects are enough to improve precision, we may do better with a non-parametric decomposition.

To motivate the decomposition, consider a model of the following form:

$$\begin{aligned} WTP_{ij} = &\overbrace{\beta_i + \sum_{k=-3}^{3} \gamma_k \mathbf{1}\{\hat{u}_j^i = k\} \times \text{Direct}_{ij}}^{\text{Self-regarding motives}} \\ &\overbrace{+ \theta \Delta \hat{a}_j^i + \sum_{k=-3}^{3} \alpha_k \mathbf{1}\{\hat{u}_j^i = k\}}^{\text{Other-regarding motives}} \\ &+ \varepsilon_{ij} \end{aligned} \tag{4}$$

[7] Specification 2 assumes that potential senders of both the Info and Pressure Messages interpret the Likert scale for $\hat{u}_j^i$ similarly and are sensitive to it in the same way. As a robustness check, I can add two more coefficients that interact the slopes on $\hat{u}_j^i$ with an indicator for the Pressure Message condition. Another robustness check is to re-run Specification 2 on the subset of recipients who are strangers.

[8] The beliefs on the effectiveness of the message ($\Delta \hat{a}_j^i$) are not incentivized for potential senders of the Info Message.

[9] The fixed-effects specification is still consistent even if β_i is arbitrarily correlated with $\hat{u}_j^i$ and $\Delta \hat{a}_j^i$, which intuitively would seem to be the case. The coefficient θ on $\Delta \hat{a}_j^i$ may also be random and could vary positively with β_i. For instance, senders with high fixed effects may also place more value on an additional registration. This idea of "random slopes" features in the structural estimation.

The Other-regarding motives on the second row capture how the sender's WTP changes in response to the predicted effect on the recipient, either in action ($\Delta\hat{a}_j^i$) or in appreciation of the message ($\hat{u}_j^i$). For instance, altruism would imply $\alpha_k > \alpha_{k-1}$. Meanwhile, the Self-regarding motives on the first row are, as the name suggests, the part of the WTP that is for the sake of the sender. For example, $\gamma_k > \gamma_{k-1}$ is the extent to which senders are sensitive to how much they think the recipient would like the message, but only when they have to send the message directly.

I assume $\alpha_0 = 0$, which sets $\hat{u}_j^i = 0$ as the base omitted category. When senders respond "Neither agree nor disagree" to whether they think the recipient would like the message, there is no altruistic cost or value to sending the message.

First, I run Specification 4, but *without* including $\Delta\hat{a}_j^i$. This provides the average WTP for each category of $\hat{u}_j^i$ and for Direct and Anonymous Messages separately, with the Other-regarding motive from the predicted effect on registration rates incorporated into the average.[10] Suppose the predicted effect of the message on registration rates is some function of $\hat{u}_j^i$ and Direct_{ij}:

$$\begin{aligned}\Delta\hat{a}_j^i &= f(\hat{u}_j^i, \text{Direct}_{ij}) \\ &= \sum_{k=-3}^{3} (P_k^1 + P_k^2 \text{Direct}_{ij}) \mathbf{1}\{\hat{u}_j^i = k\}\end{aligned}$$

Then for example, the average WTP when $\hat{u}_j^i = -1, \text{Direct}_{ij} = 1$ is:

$$WTP_{ij}(\hat{u}_j^i = -1, \text{Direct}_{ij} = 1) = \bar{\beta}_i + \gamma_{-1} + \alpha_{-1} + \theta \times (P_{-1}^1 + P_{-1}^2) \tag{5}$$

However, these parameters are not separately identified when I run Specification 4 without including $\Delta\hat{a}_j^i$. The average WTP from Equation 5 includes both Self-regarding ($\bar{\beta}_i + \gamma_{-1}$) and Other-regarding ($\alpha_{-1} + \theta \times (P_{-1}^1 + P_{-1}^2)$) motives.

Next, I run Specification 4, now *including* $\Delta\hat{a}_j^i$. This gives me the average WTP with the Other-regarding motive from $\Delta\hat{a}_j^i$ partialed out:

$$WTP_{ij}(\hat{u}_j^i = -1, \text{Direct}_{ij} = 1 | \Delta\hat{a}_j^i = 0) = \bar{\beta}_i + \gamma_{-1} + \alpha_{-1} \tag{6}$$

Now, the parameters $\bar{\beta}_i, \gamma_{-1}, \alpha_{-1}$ are separately identified. The sum of $\bar{\beta}_i$ and γ_{-1} in Equation

[10] $\beta_0 \equiv \bar{\beta}_i$ is the average WTP for an Anonymous Message with $\hat{u}_j^i = 0$.

6 estimates the Self-regarding sources of the WTP from Equation 5. Then, what remains in the average WTP from Equation 5 after subtracting the sum of $\bar{\beta}_i$ and γ_{-1} from Equation 6 is the Other-regarding cost/value of pressuring when $\hat{u}_j^i = -1$. I can do this for each level of $\hat{u}_j^i$ to decompose the WTP to send a Direct Pressure Message into Self- and Other-regarding motives.

How does this cost or value depend on the relationship between the sender and receiver (e.g., friends or strangers)? Next, I add interaction terms to Specification 3 to analyze how the Self- and Other-regarding motives depend on the sender's relationship with the recipient. In particular, I run the fixed-effects regression:

$$\begin{aligned} WTP_{ij} = \beta_i &+ \sum_f r_f \text{Reln}_{ij}^f + \left(\beta_1 + \sum_f \beta_f \text{Reln}_{ij}^f\right) \text{Direct}_{ij} \\ &+ \left(\theta + \sum_f \theta_f \text{Reln}_{ij}^f\right) \Delta\hat{a}_j^i \\ &+ (\alpha_1 + \alpha_2 \text{Direct}_{ij} + \sum_f (\alpha_{1f} + \alpha_{2f} \text{Direct}_{ij}) \text{Reln}_{ij}^f) \hat{u}_j^i \\ &+ \varepsilon_{ij} \end{aligned} \tag{7}$$

where Reln_{ij}^f is an indicator for the type of relationship between sender i and recipient j, and the omitted base category is Strangers.

How does this cost or value vary by gender and race/ethnicity on both the sender and receiver's sides? First, I analyze how the WTP varies by the sender's gender and race/ethnicity. To study the differences between gender groups, I run the pooled OLS regression:

$$\begin{aligned} WTP_{ij} = \beta_0 &+ \rho_0 \text{Female}_i + \beta_1 \text{Direct}_{ij} + \rho_1 \text{Direct}_{ij} \times \text{Female}_i \\ &+ (\theta_1 + \theta_2 \text{Female}_i) \Delta\hat{a}_j^i \\ &+ (\alpha_1 + \alpha_2 \text{Direct}_{ij} + \alpha_3 \text{Female}_i + \alpha_4 \text{Direct}_{ij} \times \text{Female}_i) \hat{u}_j^i \\ &+ \varepsilon_{ij} \end{aligned} \tag{8}$$

For example, $\rho_1 < 0, \alpha_4 > 0$ would mean that female senders have higher costs to sending Direct Pressure Messages and are more sensitive to how much they expect the recipient to

like/dislike the message.

I explore differences among race/ethnicity groups in a regression similar to Specification 8, but with indicators for race/ethnicity groups replacing the gender dummy.

Next, I also separate the sender's WTP by the *recipient's* gender and race/ethnicity. I use the same model as 8, where the demographic indicators now refer to the recipient. Furthermore, since senders provide their WTP for multiple potential recipients of varying gender and race/ethnicity, I can include sender fixed effects. These regressions reveal whether senders are more or less willing to pressure certain gender or ethnic groups.

Lastly, I consider interactions *between* the gender and race/ethnicity of the sender and the recipient. I expand Specification 8 to include dummies for each gender interaction ($M_i \times M_j, M_i \times F_j, F_i \times F_j, F_i \times M_j$). The results compare the average WTP across the interactions for both Direct and Anonymous Messages. For race/ethnicity interactions, I interact the sender's own racial/ethnic group with an indicator for whether the recipient is also in the same group. For instance, the interaction term $\text{Asian}_i \times \text{Diff. race}_{ij} = 1$ means that the sender is Asian while the recipient is not.

The specifications that include the recipient's demographic group assume that senders can infer recipients' gender and race/ethnicity from their names. Alternately, I follow the methodology in Card, DellaVigna, Funk, and Iriberri (2020) to assign each name its probability of belonging to someone who identifies as female. Then, I replace the recipient's gender indicator with this probability in the regressions. Similarly, I use the probability of recipients' names belonging to each racial/ethnic group calculated by algorithms such as NamSor.

4.2 How does receiving social pressure affect the compliance to the norm?

4.2.1 Sample

This research question focuses on the Non-registrants side. I use the 90% of Non-registrants whose info-sharing is not optional (see Section 3).

4.2.2 Primary outcome variables

The primary outcome variables are whether Non-registrants (who have not registered when they begin the study) register to vote by the election (Register_j), and if so, whether they vote (Vote_j). After the November 3 General Election, I will check the updated California voter registration file for these outcomes.

4.2.3 Secondary outcome variables (structural estimation)

In the structural estimation, I investigate the mechanisms of the message effect (if any) and the welfare cost of receiving social pressure. One potential mechanism of social pressure is to motivate action by shifting beliefs about the norm. To assess this mechanism, I look at Non-registrants' predicted final registration rate on their campuses, which is asked on both surveys. For the welfare cost of social pressure, the survey also elicits the Non-registrants' WTP to (not) share their status and potentially receive an email.

4.2.4 Covariates

- $\text{Reln}_{ji}, G_{j/i}, R_{j/i}$, and Diff. race_{ji} from Section 4.1.2.
- Received_j is an indicator for whether Non-registrant j reports receiving an email message from another participant.

4.2.5 Empirical specification

The first specification is an intention-to-treat analysis. I run the OLS regression:

$$\text{Register}_j = \kappa_0 + \kappa_1 \text{Msg}_j + S_j + \epsilon_j \tag{9}$$

with heteroskedasticity-robust standard errors. S_j is Non-registrant j's randomization stratum, and Msg_j is an indicator for Non-registrant j being assigned to receive a Direct Pressure Message. The coefficient κ_1 estimates the effect of the message on registration rates.

I augment the baseline specification by differentiating the message effect by the relationship

between the recipient and sender:

$$\text{Register}_j = \kappa_0 + \sum_f \kappa_f \text{Reln}_{ji}^f \times \text{Msg}_j + S_j + \epsilon_j \tag{10}$$

Similar to the analysis on the sender's WTP, I also separate the effects of the message by gender and race/ethnicity groups, and also explore the interactions of gender and race/ethnicity between the recipient and sender. Nevertheless, the power to detect these heterogeneous effects is small.

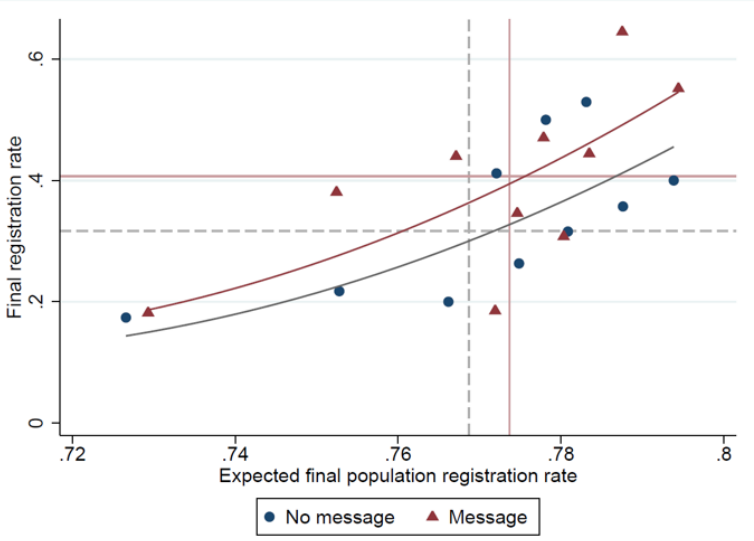
Instrumental variables analysis Some Non-registrants may miss the email in their inbox, or their assigned sender may refuse to send them a message. Given this one-sided compliance, I also estimate the treatment-on-treated/LATE effects. I replace the Msg_j indicator in Specifications 9 and 10 with the Received_j variable, and run two-stage least squares where in the first stage, Msg_j instruments for Received_j.

Structural estimation: Mechanisms and welfare analysis The mechanisms behind social pressure will be structurally estimated, but the intuition can be drawn from reduced-form results. Figure 3 plots the actual and perceived registration rates for a simulated sample of Non-registrants. In the simulation, the messaged group have higher actual and perceived registration rates. In this case, social pressure appears to work through updating perceptions of the norm. On the other hand, an increase in registration rates, but not beliefs, would indicate that social pressure alters the personal costs or benefits of registering. Lastly, changes in the slope of the curve would suggest shifts in the sensitivity to the norm.

Using an action-based social signaling model, the outcomes from scaling this social pressure intervention will be estimated under two different equilibrium assumptions. The first is the partial-equilibrium scenario in which this policy is expected, so that some prior Non-registrants will register solely from anticipating that their statuses may be revealed, but beliefs remain heterogeneous. The latter is a full equilibrium setting where all beliefs (e.g., on the norm and effectiveness of social pressure) are corrected, which may mirror long-term outcomes under repeated interactions.

The welfare analysis on the Non-registrants' side studies their WTP to (not) share their registration status. I see whether Non-registrants with stronger beliefs about the norm (i.e., higher predictions of the registration rate) are less willing to share their unregistered status. I cannot price out the cost of registering to vote since federal law prohibits offering incentives

Figure 3: Mechanisms of the message effect



for registering to vote, so I calibrate the model assuming different cost values. The results can answer policy-relevant questions such as, "What would be the welfare cost of social pressure (to both sender and receiver) for each additional registration?"

4.3 Secondary research questions

In this section, I briefly outline the analyses for the secondary research questions.

1. **Does pressuring others increase one's own chances of complying with the norm?**
   For Registrants whose assignment to send or not send a message is random,[11] I explore whether sending a social pressure message ("treatment") increases voter turnout in the election (outcome).

2. **Do people have correct beliefs about other people's compliance with the norm?**
   For each round/campus, participants make incentivized predictions about the current and final voter registration rates among participants on their campus. I compare the average predictions to the actual rates.

[11] This applies to the vast majority of Registrants. The only Registrants for whom sending is not random are those matched to a Non-registrant whose info-sharing is optional.

Registrants also make predictions on whether *individual* Non-registrants will register to vote by the election (with and without receiving a social pressure message). I investigate how well their forecasts predict the Non-registrants' post-election registration statuses.

3. **Do people know how effective giving social pressure is and how it will be received?**
   First, I compare the Registrants' average predicted effect of the social pressure message to the actual effect. Within the sample of messaged Non-registrants, I then regress their post-election registration rates on the sender's predicted effect of the message. Senders also forecast how much the recipient will like/dislike their message. I correlate their responses to the recipient's actual answers to this question. I see whether the accuracy of these predictions increases with the strength of the relationship between the sender and recipient.

4. **How do people compose social pressure messages?**
   I look at the number of words edited, deleted, and added from the provided template among those who are selected to send a message. First, as an explanatory variable, I see whether messages with more edits are more effective in getting the recipient to register. Then, as an outcome variable, I explore whether senders edit messages more when they know the recipient, or when they state a higher WTP to send the message.