Online Appendix for Experiment 1 (Unauthorized Immigration)

Joshua Kalla[*]
David Broockman[†]

Contents

[*]Yale University, Departments of Political Science and Statistics & Data Science, josh.kalla@yale.edu
[†]University of California, Berkeley, Department of Political Science, dbroockman@berkeley.edu

The full replication code and data that produces this report will be available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8BFYQO.

This experiment was pre-registered at https://egap.org/registration/5138.

Intervention Details

Training

Before beginning the experiment, the partner organizations dedicated 278 unique canvasser shifts to developing the intervention and developing the training. The canvassers in this experiment were paid. They received training when they first started, including following a more experienced canvasser for a day. Throughout the program, they received ongoing training and feedback. The trainings focused on providing canvassers with the skills to listen to and ask questions of voters in a non-judgmental manner that would elicit narratives from voters about their experiences. Trainings often involved role play and viewing video of past canvass conversations. Trainings were led by the New Conversation Initiative.

Canvasser Demographics

The canvassers for this project were primarily paid canvassers recruited by the three local organizations. Neither the canvassers nor the local organizations had prior experience conducting in-person conversations to reduce exclusionary attitudes. 77% of the conversations were conducted by canvassers age 30 and under (average age was 25), 60% by female canvassers, 54% by canvassers who self-identified as Latino, and 24% by canvassers who self-identified as immigrants.

Intervention Procedure

The canvassers were trained to follow the below procedure when approaching homes when subjects were in the treatment conditions. Being mainly concerned with external validity, this procedure does not strictly rely on only one theoretical paradigm as is common in lab studies. However, the majority of the time in the training and in the conversations was spent on how to non-judgmentally exchange narratives. (The full scripts are reproduced below.)

Canvassers themselves were not aware of the details of the experiment or the survey and nowhere in the conversation did they indicate that the effects of the conversation were being measured or part of the study.

Establish Contact

1. **Determine if voter is home.** The canvasser knocks on the door and says, “Hi, I’m [canvasser’s name]. Are you [subject’s name]?” If the subject identifies themself, the canvasser marks “Voter came to door” on their walk list. This leads the voter to be targeted for resurveying. Note that this first step is identical in the placebo and treatment conditions.

Create Non-Judgmental Context

2. **Intervention begins: inform subject about the policy being discussed.** The canvassers began the intervention by engaging in a series of strategies to elicit participants’ opinions in a non-judgmental manner. First, canvassers informed voters that they were at the door to discuss a policy related to unauthorized immigrants (e.g., in Tennessee, “Based on what you know now, would you say you are against, undecided, or in favor of large-scale arrests and detainment of undocumented immigrants at their place of work?”). Canvassers then asked voters about their opinion on the policy and then asked them to explain their position. Canvassers were trained to ask these questions in a non-judgmental

manner, not indicating they were pleased or displeased with any particular answer, but rather to appear genuinely interested in hearing the subject ruminate on the question. This was intended to encourage effortful reflection and to build rapport.

Exchange Narratives - Removed in the abbreviated condition

3. **Exchange narratives about personal experience with immigration.** The canvasser then asked the voter if they know anyone who is an immigrant and, in particular, an unauthorized immigrant. If the voter knows someone, the canvasser would have the voter talk about how they know this person, their immigration story, and how it must feel to be an immigrant. Whether or not the voter knows an immigrant, the canvasser would always share their immigration story. This might be a personal story or about a friend or family member. The canvasser would end this section by asking the voter if there is anything about the story that they can relate to, encouraging perspective taking [5].

4. **Exchange narratives about a personal experience with compassion.** The intervention attempted to prompt values that would lead participants to be more supportive of unauthorized immigrants and encourage analogic perspective-taking. To do this, canvassers asked voters to share a time when someone showed them compassion. If necessary, canvassers sometimes told their own stories of being shown compassion in order to make voters feel comfortable sharing a story of their own. For many canvassers, this would involve telling a story about being shown compassion that related to their own experiences as immigrants or as close friends or family to immigrants.

Canvassers' goal was for this non-judgmental exchange of narratives to end with individuals self-generating and explicitly stating aloud implications of the narratives that ran contrary to their previously stated exclusionary attitudes.

Address Concerns

5. **Address voter concerns**. At this point, the canvasser would return to any concerns about unauthorized immigrants that the voter may have mentioned earlier. The canvasser would talk through these concerns and, where applicable, provide talking points to refute them. Canvassers were trained not to address concerns until this point in the conversation so that voters would not feel threatened by this section. Only after rapport had been established, stories shared, and the value of compassion activated would canvassers address concerns. Canvassers would address concerns at this point in both the Full and Abbreviated interventions.

Make the Case

6. **Provide arguments and information**. The canvasser would then reiterate for the voter why they were canvassing and why they hoped the voter would become more supportive of unauthorized immigrants.

Encourage Active Processing

7. **Ask for opinion again; rehearse opinion change.** The intervention ended with canvassers asking voters if and why the conversation changed their exclusionary attitudes towards unauthorized immigrants. Rehearsal of opinion change is a strategy that has been shown to facilitate active processing and increase the persistence of attitude changes [13]. The canvasser then thanked the subject and left.

Placebo Procedure

The sole purpose of the placebo conversations was to identify voters who were home and thus voters with whom the intervention could be plausibly attempted (see section entitled "Placebo Design"). When approaching homes where subjects were in the placebo group, canvassers followed the following procedure instead:

1. **Determine if voter is home.** The canvasser knocks on the door and says, “Hi, I’m [canvasser’s name]. Are you [subject’s name]?” If the subject identifies themself, the canvasser marks “Voter came to door” on their walk list. This leads the voter to be targeted for resurveying. Note that this first step is identical in the placebo and treatment conditions.
2. **Placebo begins.** The canvasser would then deliver the placebo survey, which varied by site.
3. **Conversation ends.** The canvasser thanks the subject and leaves.

Scripts

Below are the scripts for the Full Intervention, Abbreviated Intervention, and Placebo conditions.

Start Time: ______ **CONVO TYPE: Deep Imm**

Introduction

Hi, are you_____? *[If not the person on list]* I'm talking to voters in the neighborhood about some issues that have been in the news lately. May I speak with ____?

Great! I'm _____, a volunteer with Welcoming Tennessee Initiative and we're talking to voters in your neighborhood about undocumented immigrants and the American way of life.

Based on what you know now, would you say you are against, undecided, or in favor of large-scale arrests and detainment of undocumented immigrants at their place of work?

Step 1: Uncover Voter's Honest Opinions

To be more specific, on this 0-10 scale, where 0 is you're absolutely in favor of large-scale workplace raids, and 10 is you're absolutely against large-scale workplace raids, where would you put yourself?

0	1	2	3	4	5	6	7	8	9	10
In Favor					**<<<Undecided>>>**					**Against**

Why is that the right number for you?

[If less than 10] What is on either side of the issue for you? What are some reasons that you would be in favor? Against?

Step 2: Connect Around Experiences w/ Immigration

Again, my name is ______ and we are out here having open/honest conversations about immigration because it's not something we do very often.

What is your experience with immigration? How did that feel?

Do you know anyone who has immigrated to the US? □ NO □ YES
Do you know any undocumented immigrants?

<u>No- *Doesn't Know Someone*</u>	**<u>Yes- *They Know Someone*</u>**
• Maybe an acquaintance? A friend's parents? A coworker? • What kind of role do you see immigrants playing in your community? • ***Share your immigration story***	• **Who are you closest to?** May I ask, how they are doing? • **What is their story? What is their documentation status?** • What do you think that was like for them? Tell me more? What is their name?

<u>SHARE *YOUR* IMMIGRATION STORY**</u>**

I'm out here today because **I**... **My family**….**My friend_____**

[If voter did not share about someone undocumented] **Do you know anyone who has a story like that? Is there anything about my story you can relate to?**

Step 3: Connect Around Compassion Experiences

I think having these conversations is important because it gives us a chance to think about **how we want to treat everyone in our community, including undocumented immigrants**, because we've all faced tough times and needed others…

 End Time: _____

** Your Compassion Story **	**Voter's Compassion Story**
I remember when… *Go slow…paint a picture* *Talking about the feeling*	**When was a time when someone showed you compassion when you really needed it**? Maybe a teacher…friend…parent? What was the situation? How old were you? How did that feel? Why?

Step 4: Address Concerns

Thanks so much for having this conversation with me… Earlier you mentioned______ as a concern? What are your fears? What is on your mind now? What are you picturing might happen? Do you have a personal connection to that concern?

- Would you like to see the law changed? Are there times in history where you think we've gotten the law wrong?
- Most immigrants that are undocumented would also like to see the law changed. Given the opportunity, we/they would like to _______.
- In the meantime, what should we do? How should we treat people? Do we want people living in fear/uncertainty/instability?

[Legality] Our immigration system is broken. For most, it can take over 20 years to legally immigrate to the U.S. …many are fleeing violence and poverty and cannot wait that long.

[Criminality] I think because not everyone knows someone undocumented, people have concerns about increased crime. Research shows that across the country, cities with the most immigrants actually have lower crime rates than 40 years ago.

[**Workplace raids]** The first large-scale worksite raid in a decade happened just a few hours from here near Morristown, Tennessee. In April, ICE arrested 97 workers at a meat-packing plant, some of whom had lived in the community for decades and had put in several years of honest labor at the factory. The day after the raid, nearly 600 kids from one rural school district stayed home from school out of fear.

Ten years ago our federal government stopped using mass worksite raids as a way to enforce immigration law. These militaristic operations separated hundreds of families, terrorized immigrant communities, and devastated local economies.

Step 5: Make Your Case

I think it's important to end workplace raids because I want ***everyone*** in our community, **including undocumented immigrants**, *my* _____ *and your* _____, to be treated with compassion and not to live in fear and instability.
[Explain why it is important to you personally]

Step 6: FINAL RATING

Thank you, again, for your time. To wrap things up, back on the 0-10 scale, where 0 is you're absolutely in favor of large-scale workplace arrests, and 10 is you're absolutely against large-scale workplace arrests, and 5 is your undecided, where would you put yourself?

0	1	2	3	4	5	6	7	8	9	10
In Favor					**<<<Undecided>>>**					**Against**

[If moved] Why is that the right number? What makes you rate yourself differently?

Start Time: ______ **CONVO TYPE: Short Imm**

Introduction

Hi, are you_____? *[If not the person on list]* I'm talking to voters in the neighborhood about some issues that have been in the news lately. May I speak with ____?

Great! I'm _____, a volunteer with Welcoming Tennessee Initiative and we're talking to voters in your neighborhood about undocumented immigrants and the American way of life.

Based on what you know now, would you say you are against, undecided, or in favor of large-scale arrests and detainment of undocumented immigrants at their place of work?

Step 1: Uncover Voter's Honest Opinions

To be more specific, on this 0-10 scale, where 0 is you're absolutely in favor of large-scale workplace raids, and 10 is you're absolutely against large-scale workplace raids, where would you put yourself?

0	1	2	3	4	5	6	7	8	9	10
In Favor					**<<<Undecided>>>**					**Against**

Why is that the right number for you?

[If less than 10] What is on either side of the issue for you? What are some reasons that you would be in favor? Against?

Step 4: Address Concerns

Thank you for sharing [Choose 1]

[Legality] Our immigration system is broken. For most, it can take over 20 years to legally immigrate to the U.S. …many are fleeing violence and poverty and cannot wait that long.

[Criminality] I think because not everyone knows someone undocumented, people have concerns about increased crime. Research shows that across the country, cities with the most immigrants actually have lower crime rates than 40 years ago.

[Workplace raids] The first large-scale worksite raid in a decade happened just a few hours from here near Morristown, Tennessee. In April, ICE arrested 97 workers at a meat-packing plant, some of whom had lived in the community for decades and had put in several years of honest labor at the factory. The day after the raid, nearly 600 kids from one rural school district stayed home from school out of fear.

Ten years ago our federal government stopped using mass worksite raids as a way to enforce immigration law. These militaristic operations separated hundreds of families, terrorized immigrant communities, and devastated local economies.

Step 5: Make Your Case

I think it's important to end workplace raids because I want ***everyone*** in our community, **including undocumented immigrants**, to be treated with compassion and not to live in fear and instability.

Step 6: FINAL RATING

Thank you, again, for your time. To wrap things up, back on the 0-10 scale, where 0 is you're absolutely in favor of large-scale workplace arrests, and 10 is you're absolutely against large-scale workplace arrests, and 5 is your undecided, where would you put yourself?

0	1	2	3	4	5	6	7	8	9	10
In Favor					**<<<Undecided>>>**					**Against**

End Time: _____

[If moved] Why is that the right number? What makes you rate yourself differently?

Start Time: ______ **CONVO TYPE: News Surv**

News Survey

Hi, are you_____? *[If not the person on list]* I'm talking to voters in the neighborhood about some issues that have been in the news lately. May I speak with ____ ?

Great! My name is ______ and I am with the Welcoming Tennessee Initiative. We're out talking with voters in the neighborhood today about how they stay informed. We have one quick question about the news.

When you think about where you get most of your news, would you say that it mostly comes from Local TV, Cable TV, Radio, Internet, Print Newspaper, Word of Mouth or someplace else?

[*Record response*: Local TV, Cable TV, Radio, Internet, Print Newspaper, Word of Mouth or someplace else]

Thank you!

We will be using this information to better understand how to reach Tennesseans on important issues.

End Time: ______

Survey Recruitment Procedures and Experimental Design

In this section we describe the survey recruitment procedures and the experimental design. We assess the representativeness of the sample at each step and test design assumptions in other sections.

Baseline Survey

To measure the effects of the intervention, we conducted ostensibly unrelated surveys of voters living in two regions of California and one region of Tennessee. To recruit voters to these surveys, the partner organizations first provided us with contact information for voters living in the areas they planned to canvass, acquired from the publicly available list of registered voters. We invited these voters to the baseline survey by mail. The survey was called the "UNIVERSITY-UNIVERSITY Public Opinion Study". (See more detail in "Additional Survey Details" section.)

The recruitment letter included the survey web URL, a unique login for each voter, and instructions for taking the survey online. To participate, respondents entered the URL from the letter in their computer or smartphone and then their unique login. We mailed letters to 122712 households that contained individual logins for 217600 people; when multiple eligible voters lived in the same household, we sent the household one letter that contained a unique login for each person.

Voters were offered no incentives for completing the baseline survey but were offered $5 for completing each follow-up survey. Voters received these incentives via email (collected during the survey) immediately upon completion of the survey. Voters could redeem these $5 incentives as gift-cards to Amazon, iTunes, Starbucks, WalMart, or Home Depot or as donations to Habitat for Humanity, the National Parks Foundation, or Clean Water Fund.

Random Assignment of Households

7870 voters completed the baseline survey and provided a valid email address. We randomly assigned one-third to each treatment condition. Voters were randomly assigned at the household level, ensuring that multiple voters who completed the pre-survey within the same household were always assigned to the same treatment condition. All analyses adjust standard errors to account for this clustered assignment (see details below). This procedure is identical to that used in [2] and follows the best practices for field experiments with survey outcomes [3].

The household-level clustered random assignment took place within blocks of three households. These blocks were formed by matching households on household size and on household-level-average values of baseline covariates (a factor measuring baseline views on immigration and a factor measuring baseline views on partisan politics). Within each block, one household (cluster) was assigned to each condition. This pre-treatment blocking reduces the chance of imbalance between conditions and improves precision [3].

Random Assignment of Turfs

On the day of each canvass, groups of households were formed into "turfs" by the staff at the partner organizations. "Turfs" are groups of nearby households convenient for two canvassers to visit by walking a short distance. Households were put in groups blind to treatment assignment and simply based on the geographic layout of households to be canvassed that day. A route connecting the households in the turf were then drawn, again blind to treatment assignment, such that an efficient route could be followed; half of the households were marked for a Canvasser A and half for a Canvasser B in an order each canvasser could follow. The groups of households (turf) were then randomly assigned to pairs of canvassers by having canvassers pick a number corresponding to a turf out of a hat. Then, canvass leaders flipped a coin to determine which canvasser would knock on A doors and which on B doors. In some cases, Canvasser A and B would be one person. Data-quality checks conducted after the canvass ensured that canvassers all properly canvassed the assigned doors within their turf. This random assignment of canvassers to turf allows us to assess canvasser-level treatment effect heterogeneity, such as by canvasser immigration status.

Placebo Design for Delivering Intervention

Canvassers attempted to have a conversation about unrelated issues with voters in the placebo group and a conversation containing the intervention with voters in the two treatment groups. In Fresno, the placebo was to get voters to sign a petition on gun violence; in Orange County the placebo was a brief survey on rent control; in Tennessee, the placebo was a brief survey on media consumption. This placebo-controlled experimental design [12] is common in studies of door-to-door canvassing interventions and field experiments more generally [3]. Nickerson [12] summarizes the placebo design:

> Rather than rely upon a control group that receives no attempted treatment, the group receiving the placebo can serve as the baseline for comparison for the treatment group. . . assuming that (1) the two treatments have identical compliance profiles; (2) the placebo does not affect the dependent variable; and (3) the same type of person drops out of the experiment for the two groups.

Gerber et al. [7] similarly summarize the design:

> subjects who agree to participate in a study and for whom the prospect of treatment is imminent are randomly assigned to receive either the treatment or the placebo.

The sole purpose of the placebo discussion was to identify subjects who were home and thus with whom a conversation at the door could be attempted (versus subjects who were not home at all or would not even open the door). Identifying this group allows a direct comparison of subjects with whom the intervention actually began to subjects with whom the intervention could have begun but did not because of their random assignment (and thus with whom a conversation about recycling began instead). This design dramatically improves the precision of door-to-door canvassing experiments [3].

We implemented the placebo design as follows.

First, the canvassers began by implementing an identical procedure regardless of experimental condition. Canvassers were given walk lists of voters to contact that had been sequentially ordered by voters' addresses blind to voters' treatment assignment. Canvassers proceeded down the list of houses in the experiment in this order, knocking on one door after another without regard to the household's experimental group. The beginning of the conversation was also identical in each condition: "Hi, are you [subject's name]?" If the subject identified him/herself or came to the door at this point, the canvasser then checked a box called "Voter came to door" on their walk list. The experimental sample consists of those who came to the door at this point.

Only after canvassers determined whether the voter they were looking for came to the door or not did they begin either implementing the intervention or delivering the placebo. Importantly, nothing was different in the procedure before this point: voters did not know the canvasser intended to have a conversation with them about immigration or the placebo issues before identifying themselves or not; canvassers did not inform voters about the topic of the conversation before this point.

These procedures guarantee an unbiased experimental comparison among voters who came to the door and then were delivered the intervention or were then not delivered the intervention based on their random assignment [12, 3].

One strength of this study's research design is that we are able to sensitively test the placebo design's key assumption: the kinds of voters who identify themselves at their doors before the placebo starts and before the intervention starts are similar. Our tests support this assumption. We describe these tests in the *Tests of Design Assumptions* section.

Follow-Up Surveys

Following the placebo design described above, we conducted multiple waves of follow-up surveys for voters who came to the door in any condition. These follow-up surveys began around 1 week, 1 month, and 3-6 months after the day each voter was canvassed. We solicited voters to complete these re-surveys at the email addresses they provided in the baseline survey. Three reminders to complete the follow-up surveys were sent for each survey wave.

Note that to the extent any voters answered the wrong surveys or did not answer the surveys carefully, this measurement error would lead us to underestimate the true effects of treatment [6].

Additional Survey Details

The survey was called the UNIVERSITY-UNIVERSITY Opinion Study, conducted by University #1 and University #2. The survey was conducted by the authors using a panel initially recruited through the mail and then managed using Qualtrics via e-mail, using the e-mail addresses subjects provided us.

The population refers to registered voters in selected neighborhoods in Fresno, Orange County, and Tennessee, as chosen by staff at the partner organizations. Voters were recruited from this population by mail we sent to their household.

The below table shows how the representativeness of those who responded to the survey differ from those mailed an invitation to participate in the survey. These data come from the voter file. Note that no weighting is used in the analysis; the aim of the estimation is to test for the existence of treatment effects within this sample, not to generalize to the population of invited respondents.

Table OA1: Representativeness of Experiment at Each Stage

Sample	Female	Age	AfAm	Latino	Voted 16	Voted 14	Voted 12	Voted 10	Voted 08	N
Starting	0.51	49.08	0.05	0.16	0.74	0.4	0.63	0.45	0.59	217600
Baseline Resp.	0.52	50.03	0.03	0.11	0.85	0.55	0.71	0.55	0.66	7870
Canvassed	0.51	52.23	0.02	0.1	0.89	0.6	0.75	0.61	0.71	2374
1 Wk Resp.	0.52	52.38	0.02	0.09	0.9	0.63	0.75	0.63	0.72	1578
1 Mo Resp.	0.52	52.46	0.02	0.09	0.91	0.63	0.76	0.63	0.73	1508
3-6 Mo Resp.	0.51	52.49	0.02	0.1	0.91	0.63	0.77	0.64	0.73	1384

Outcomes

The survey included dozens of political, social, and cultural questions, only some of which were related to immigration. In our pre-analysis plan, we indicate which items constituted experimental outcomes. Below we list these items and give their full text.

The below items appeared on multiple surveys; the # sign below will be replaced with the survey number in our analysis:

- The baseline survey is survey 0;
- the 1-week survey is survey 1;
- the 1-month survey is survey 2, and;
- the 3-6 month survey is survey 3.

The variable name for each item is written using `in-line code`. For the remainder of the paper, we will refer to these items by their variable names.

Anti-Immigrant Prejudice Index

The first set of questions are five point scales where respondents were asked: "Do you agree or disagree with the below statements about undocumented or illegal immigrants?" Response options were: Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree:

- `t#_imm_prej_living`: "I would have no problem living in areas where undocumented immigrants live."
- `t#_imm_prej_fit`: "Too many undocumented immigrants just don't want to fit into American society."
- `t#_imm_prej_burden`: "Undocumented immigrants are too much of a burden on our communities."
- `t#_imm_prej_crime`: "Undocumented immigrants have already broken the law coming here illegally, so they are more likely to commit other crimes."
- `t#_imm_prej_values`: "Undocumented immigrants hold the same values as me and my family."

Respondents were also asked a feeling thermometer:

- `t#_therm_illegal_immigrant`: Feeling thermometer towards "illegal immigrants".

Respondents were asked a variant of the Bogardus social distance scale question [1]. They were coded as 1 if they answered Relative; 2 for Friend; 3 for Neighbor; 4 for Coworker; 5 for Resident of My State; 6 for None:

- `t#_social_distance_immigrant`: "Below are some groups of people. Look at each of them and say which is the closest relationship you would find acceptable for each group. For example, if you would accept someone from a group living on your street, but not as a close friend, then you would choose neighbors... Undocumented Immigrant". (Note that due to a coding error, this was not included in the t3 survey.)

Anti-Immigrant Policy Index

Respondents were first asked: "Politicians are considering a number of policies about immigration. We want to know what you think. Do you agree or disagree with the statements below?" Response options were: Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree:

- `t#_imm_attorney`: "The government should provide legal aid to all undocumented immigrants who cannot afford their own attorney for legal or courtroom deportation proceedings." This was the question specific to Fresno.
- `t#_imm_police`: "Local police should ask for documentation and automatically turn immigrants over to federal immigration officers when they are found to be in the country illegally." This was the question specific to Orange County.
- `t#_imm_deportall`: "The federal government should work to identify and deport all illegal immigrants, including in the workplace." This was the question specific to Tennessee.
- `t#_imm_daca`: "The federal government should grant legal status to people who were brought to the US illegally as children and who have graduated from a U.S. high school."
- `t#_imm_citizenship`: "The federal government should allow undocumented immigrants currently in the U.S. to become citizens after they have lived, worked, and paid taxes for at least 5 years."
- `t#_imm_compassion`: "Undocumented immigrants deserve compassion and should not live in daily fear of deportation."

Perspective Taking Index

In addition to reducing exclusionary attitudes, we were also interested in measuring whether the canvass increased respondents' abilities to take the perspectives of undocumented immigrants. This was not the primary purpose of the intervention.

Respondents were asked: "Do you agree or disagree with the below statements about undocumented or illegal immigrants?" Response options were: Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree:

- `t#_imm_persp_imagine`: "I can imagine how things look from undocumented immigrants' perspective." This question is abbreviated from [10].
- `t#_imm_persp_difficult`: "I find it difficult to see things from an undocumented immigrants' point of view." This question is abbreviated from [4].

Active Processing Index

We asked the same on respondents' ability to actively process on immigration, which similarly was not the primary purpose of the intervention.

- `t#_imm_actproc_thought`: "I have thought a lot about how we should treat undocumented immigrants in our community."
- `t#_imm_actproc_confident`: "I feel confident in my ability to distinguish good from bad immigration policies." This question is abbreviated from [9].

Outcome Indices

In our pre-analysis plan, we specified that we would combine multiple items into indices to test hypotheses. Combining outcomes into an index increases precision by decreasing survey measurement error and limits the potential for multiple hypothesis testing [3].

The indices, to be described momentarily, are as follows:

- `t#_factor_prej`: An index of outcomes from t#__factor__overall capturing the Anti-Immigrant Prejudice Index.
- `t#_factor_policy`: An index of outcomes from t#__factor__overall capturing the Anti-Immigrant Prejudice Index.
- `t#_factor_overall`: An index of all primary outcomes (i.e., all the items in the prejudice and policy indicies), created to test the omnibus hypothesis that the treatment had any effects.
- `t#_factor_persptake`: An index of outcomes from the Perspective Taking Index.
- `t#_factor_actproc`: An index of outcomes from the Active Processing Index.

In addition to an outcome index at each time period, we also present the results of a pooled outcome index — `tALL_factor_` — taking the average of each outcome index across time periods. If a respondent did not take a particular post-treatment survey, that time period is excluded from the average. Note that this pooled outcome was not pre-specified in our pre-analysis plan. We use this pooled outcome index to increase the precision of our treatment effect estimates through further reductions in measurement error. This pooled outcome is also a useful brief summary of the overall effects. We calculate the pooled outcome using the below Stata code:

```
// Generate factor averages for pooling
foreach factor in prej policy overall actproc persptake {
    egen tALL_factor_`factor' = rowmean(t1_factor_`factor' t2_factor_`factor' t3_factor_`factor')
}
```

Procedure for Combining Outcomes into Indices

We pre-specified that we would create the indices by using factor analysis and rescaling the factors to have mean 0 and standard deviation 1.

We use the below Stata code to generate the factors. Note that we code all indices such that higher values on the indices indicate more tolerance and success of the intervention. If a factor is reverse-coded, we multiply by -1 to adjust for this.

```
factor [VARIABLES USED], fa(1)
predict t#_[FACTOR NAME]_temp
```

```
egen t#_[FACTOR NAME] = std(t#_[FACTOR NAME]_temp) // standardize to mean 0, SD 1
```

Estimation Procedures

Average Treatment Effects

Consistent with our pre-analysis plan, to estimate treatment effects we use ordinary least squares (OLS) regressions with cluster-robust standard errors, clustering on household and also including the pre-treatment covariates from the baseline survey and voter list named in our pre-analysis plan. This procedure and these covariates were pre-specified in advance and produce unbiased estimates of causal effects [6, 3]. Note that there is no reclassification of treatment based on what occurs at the door and we do not exclude any subjects who came to the door; we compare all subjects who came to the door and were pre-assigned to the treatment conversation to all subjects who came to the door and were pre-assigned to the placebo conversation.

Contact Rate

Contact is defined as the voter coming to the door and being identified before the topics of the placebo or immigration begin. Across the three conditions among voters who responded to the baseline survey and were then randomly assigned to an experimental condition, the contact rates were:

- Placebo: 0.31.
- Abbreviated Intervention: 0.29.
- Full Intervention: 0.31.

In the two intervention conditions, we asked the voters to first answer a rating question about immigration. The share of voters who answered this question provides an indication of how often voters actually began the conversation. Conditional on having come to the door, the proportion of voters who provided this first rating is:

- Abbreviated Intervention: 0.73.
- Full Intervention: 0.68.

Across the three conditions, among the voters who came to the door, the canvass completion rates were:

- Placebo: 0.86.
- Abbreviated Intervention: 0.76.
- Full Intervention: 0.66.

Tests of Design Assumptions

Covariate Balance among All Subjects, Compliers, and Reporters

The below tables demonstrate that balance on pre-treatment observable attributes is maintained among the original universe of pre-survey respondents randomized to each group, the sub-sample that was canvassed, and the sub-sample that was both canvassed and successfully re-interviewed. Each table shows the mean value for the covariate under each condition as well as the p-value from a one-way ANOVA test. The first table considers all voters who were randomly assigned after having taken the pre-survey (all subjects); the second table considers all voters who were successfully contacted (compliers); the remaining tables consider all voters who responded to the first through third post-surveys (reporters).

Table OA2: Covariate Balance among Pre-Survey Respondents.

	Placebo	Abbrev Intervention	Full Intervention	p-value
Age	50.02	50.04	50.02	1
Female	0.51	0.52	0.52	0.87
Latino	0.11	0.11	0.11	0.82
Legal Immigrant Feeling Thermometer t0	82.95	83.55	82.44	0.15
Illegal Immigrant Feeling Thermometer t0	47.16	47.23	47.52	0.89
Baseline Factor of Support	0.00	0.00	0.00	0.99
N	2623.00	2623.00	2624.00	-

Table OA3: Covariate Balance among Compliers.

	Placebo	Abbrev Intervention	Full Intervention	p-value
Age	52.07	52.09	52.51	0.83
Female	0.51	0.52	0.50	0.76
Latino	0.12	0.09	0.11	0.16
Legal Immigrant Feeling Thermometer t0	82.43	83.45	82.64	0.6
Illegal Immigrant Feeling Thermometer t0	46.97	47.82	47.40	0.84
Baseline Factor of Support	-0.03	0.03	0.01	0.43
N	814.00	748.00	812.00	-

Table OA4: Covariate Balance among 1st Post-Survey Respondents.

	Placebo	Abbrev Intervention	Full Intervention	p-value
Age	51.42	52.76	52.98	0.25
Female	0.52	0.52	0.53	0.89
Latino	0.11	0.08	0.09	0.16
Legal Immigrant Feeling Thermometer t0	84.06	84.40	82.96	0.48
Illegal Immigrant Feeling Thermometer t0	48.43	48.93	48.66	0.96
Baseline Factor of Support	0.05	0.11	0.08	0.62
N	536.00	499.00	543.00	-

Table OA5: Covariate Balance among 2nd Post-Survey Respondents.

	Placebo	Abbrev Intervention	Full Intervention	p-value
Age	51.48	53.17	52.79	0.23
Female	0.52	0.50	0.52	0.84
Latino	0.09	0.08	0.10	0.64
Legal Immigrant Feeling Thermometer t0	84.33	84.67	83.03	0.4
Illegal Immigrant Feeling Thermometer t0	48.38	49.27	49.55	0.79
Baseline Factor of Support	0.06	0.12	0.09	0.56
N	514.00	486.00	508.00	-

Table OA6: Covariate Balance among 3rd Post-Survey Respondents.

	Placebo	Abbrev Intervention	Full Intervention	p-value
Age	51.83	52.61	53.00	0.53
Female	0.50	0.49	0.54	0.31
Latino	0.11	0.09	0.10	0.6
Legal Immigrant Feeling Thermometer t0	84.15	85.18	82.49	0.12
Illegal Immigrant Feeling Thermometer t0	48.25	48.47	48.79	0.96
Baseline Factor of Support	0.05	0.09	0.09	0.75
N	459.00	448.00	477.00	-

Survey Attrition

An important design assumption is that the treatment does not affect the composition of the individuals who take each follow-up survey [3]. We investigate this by regressing an indicator for responding to a post-treatment survey on indicators of treatment assignment. Across the three survey waves, we find no evidence of differential attrition.

Table OA7: Test for differential attrition

	Effect	SE	t.stat	p
1 Week				
Abbrev	-0.01	0.01	-1.28	0.20
Full	0.00	0.01	0.23	0.82
1 Month				
Abbrev	-0.01	0.01	-0.98	0.33
Full	0.00	0.01	-0.22	0.83
3-6 Months				
Abbrev	0.00	0.01	-0.40	0.69
Full	0.01	0.01	0.65	0.52

Test of Differential Attrition by Covariates

The above subsection demonstrated that there was no average differential attrition; now, we test for whether the treatment caused attrition to differ by covariates (for example, whether it encouraged already-supportive subjects to complete the post-survey but also discouraged unsupportive subjects from doing so) [6]. To test whether attrition patterns are similar by covariates in treatment and placebo, we use a linear regression of whether or not an individual responded to the follow-up survey on treatment, baseline covariates, and treatment-covariate interactions. We then perform a heteroskedasticity-robust F-test of the hypothesis that all the interaction coefficients are zero. This procedure was pre-specified in our pre-analysis plan and is standard practice [6]. Below we report the p-value of this F-test. Based on the results presented in the Table below, there does not appear to be evidence of asymmetrical attrition.

Table OA8: p-value by Survey Wave Test of Differential Attrition by Covariates.

1 Week Survey (t1)	0.57
1 Month Survey (t2)	0.85
3-6 Month Survey (t3)	0.6

Results

Below we report the results in tabular form at each time period and for each outcome measure. In each section, the first table shows the results by the average treatment effect. Each table includes two models: one in which we adjust for the pre-specified pre-treatment covariates to improve precision and a second unadjusted model.

Overall Index of Exclusionary Attitudes

Below we present the ATE on the overall index. Note that we pre-registered a focus on the estimates with covariates (which were also pre-registered) since we expected these to be much more precise; the experimental design was intended to draw significant statistical power from the baseline survey. However, we also present results without covariates for completeness.

Table OA9: ATE effects on overall index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Full vs. Placebo	0.090	0.024	3.792	0.000	0.112	0.064	1.764	0.078
Abbrev. vs. Placebo	0.016	0.023	0.664	0.507	0.077	0.064	1.192	0.233
Full vs. Abbrev.	0.073	0.024	3.110	0.002	0.036	0.064	0.557	0.578
1 Month								
Full vs. Placebo	0.068	0.024	2.774	0.006	0.110	0.066	1.687	0.092
Abbrev. vs. Placebo	0.035	0.024	1.487	0.137	0.106	0.065	1.645	0.100
Full vs. Abbrev.	0.033	0.025	1.326	0.185	0.004	0.066	0.061	0.952
3-6 Months								
Full vs. Placebo	0.070	0.027	2.612	0.009	0.115	0.068	1.685	0.092
Abbrev. vs. Placebo	0.018	0.026	0.696	0.487	0.072	0.069	1.040	0.298
Full vs. Abbrev.	0.052	0.026	2.013	0.044	0.043	0.068	0.629	0.529
Pooled								
Full vs. Placebo	0.082	0.021	3.887	0.000	0.093	0.060	1.564	0.118
Abbrev. vs. Placebo	0.022	0.020	1.097	0.273	0.072	0.060	1.200	0.230
Full vs. Abbrev.	0.059	0.021	2.822	0.005	0.021	0.060	0.344	0.731

Prejudice Index

Below we present the ATE on the prejudice index.

Table OA10: ATE effects on prejudice index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Full vs. Placebo	0.068	0.027	2.469	0.014	0.092	0.063	1.445	0.149
Abbrev. vs. Placebo	0.004	0.027	0.146	0.884	0.054	0.065	0.825	0.409
Full vs. Abbrev.	0.064	0.027	2.396	0.017	0.038	0.063	0.598	0.550
1 Month								
Full vs. Placebo	0.066	0.028	2.357	0.019	0.109	0.065	1.674	0.094
Abbrev. vs. Placebo	0.040	0.027	1.498	0.134	0.100	0.065	1.537	0.124
Full vs. Abbrev.	0.026	0.028	0.929	0.353	0.009	0.066	0.130	0.897
3-6 Months								
Full vs. Placebo	0.053	0.030	1.769	0.077	0.101	0.068	1.495	0.135
Abbrev. vs. Placebo	0.022	0.030	0.722	0.470	0.074	0.069	1.074	0.283
Full vs. Abbrev.	0.033	0.030	1.107	0.268	0.027	0.068	0.397	0.691
Pooled								
Full vs. Placebo	0.072	0.024	3.020	0.003	0.084	0.059	1.414	0.157
Abbrev. vs. Placebo	0.016	0.023	0.695	0.487	0.058	0.060	0.967	0.334
Full vs. Abbrev.	0.055	0.024	2.355	0.019	0.025	0.060	0.417	0.677

Policy Index

Below we present the ATE on the policy index.

Table OA11: ATE effects on policy index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Full vs. Placebo	0.106	0.026	4.122	0.000	0.126	0.064	1.960	0.050
Abbrev. vs. Placebo	0.026	0.026	1.016	0.310	0.096	0.064	1.512	0.131
Full vs. Abbrev.	0.078	0.026	2.947	0.003	0.029	0.064	0.457	0.648
1 Month								
Full vs. Placebo	0.064	0.027	2.393	0.017	0.104	0.066	1.571	0.116
Abbrev. vs. Placebo	0.027	0.026	1.029	0.304	0.106	0.064	1.653	0.098
Full vs. Abbrev.	0.038	0.027	1.394	0.164	-0.002	0.066	-0.035	0.972
3-6 Months								
Full vs. Placebo	0.082	0.030	2.782	0.006	0.122	0.069	1.770	0.077
Abbrev. vs. Placebo	0.013	0.029	0.442	0.659	0.065	0.070	0.936	0.349
Full vs. Abbrev.	0.069	0.028	2.448	0.015	0.056	0.068	0.836	0.403
Pooled								
Full vs. Placebo	0.088	0.023	3.885	0.000	0.097	0.060	1.618	0.106
Abbrev. vs. Placebo	0.029	0.022	1.282	0.200	0.084	0.060	1.401	0.162
Full vs. Abbrev.	0.058	0.022	2.570	0.010	0.013	0.060	0.217	0.828

Perspective Taking and Active Processing Index

Per the pre-analysis plan, as exploratory hypotheses we also asked a perspective-taking index and an active processing-index. Our pooled results are consistent (but not statistically significant) with the full intervention producing more perspective taking than the placebo and abbreviated intervention. Similarly, our results are consistent (but not statistically significant) with both interventions potentially increasing participants' abilities to think actively about immigration.

With this said, for perspective-taking, upon reflection, the results are difficult to interpret because it is unclear in what direction we might expect effects. For example, a participant who was swayed by the intervention and became less prejudiced towards unauthorized immigrants might say they have difficulty seeing things from an undocumented immigrant's point of view because they now have a much better understanding of the challenges undocumented immigrants face.

Future research should further investigate these mechanisms.

Complier Average Causal Effects

In the pre-analysis plan, we noted that we planned to adjust for the complier average causal effect (CACE) by dividing the average treatment effect (ATE) by the proportion of conversations where the voter answered the first rating question. This CACE assumes that 1) there was no effect of the intervention for the voters who immediately refused to talk, and 2) there are no defiers; that is, no voters only received the intervention if they were assigned to the placebo group yet would not have received it were they actually in the treatment group [6]. Reporting these point estimates would not change the experimental comparison we conduct, but would increase point estimates to account for the measurement error in the treatment indicator.

Although they would be larger, we do not focus on the CACE estimates for two reasons. First, there are multiple reasonable ways to define compliance in this setting. Subjects ended some of the conversations moments after they identified themselves at the door (and hence are included in our sample); others continued further into the conversation but ended the interaction before it was completed. Compliance in this setting is therefore inherently a continuum, and there is no point in the conversation where subjects go from fully non-compliant to fully compliant. Second, there are slightly different compliance rates in the Full and Abbreviated interventions, simply by chance. This complicates comparisons of the Full and Abbreviated interventions when adjusting for compliance. (The Abbreviated Intervention actually had a slightly higher compliance rate, meaning such an adjustment would make our results about the differences between the conditions stronger.)

At the same time, we understand some readers will be interested in a CACE. To compute a CACE, we use a conservative definition of compliance, whether subjects got to the "first rating" part of the conversation where they initially told canvassers how they felt about the policy. Using this definition of compliance, the compliance rate in the Full Intervention condition was 68% and the contact rate in the Abbreviated Intervention condition was 73%. The ITT point estimates should therefore be divided by 1.48 and 1.36 in the Full and Abbreviated interventions, respectively, to compute the CACE. That is, given a true effect on compliers of 1.48 and 1.36 times the size of the ATEs we observed, we would on average estimate ATEs of the magnitude that we did. These imply CACE estimates in the Full and Abbreviated intervention conditions of, respectively, d = 0.122 and d = 0.031.

Heterogeneous Treatment Effects

In this section we present heterogeneous treatment effects by pre-specified subgroups. For each result, we present the conditional average treatment effect, adjusting for covariates, within each subgroup. As the outcome in these analyses we use the index of all the items in the prejudice and policy indicies as measured in the first post-treatment survey.

We pre-specified that we would investigate heterogeneous treatment effects by voter and canvasser traits. We will investigate these by comparing the ATEs within the pre-specified subgroups. The primary groups are:

- By canvasser immigration status: Answered "yes" to "Do you consider yourself to be an immigrant?"
- By t0 having a close friend, colleague, or family member who is undocumented in the baseline survey (yes vs. all other responses).
- By t0 being born outside of the US.

The more exploratory groups are:

- By site.
- By canvasser race: Latino-only vs. White-only vs. Other.
- By canvasser age: 30 and under vs. over 30.
- By t0 party: Democrat vs. Republican vs. Independent/Other.
- By t0 race: Latino-only vs. White-only vs. Asian-only vs. African American-only vs. Other.
- By t0 economic status: Excellent/good personal financial situation vs. Other.
- By t0 education: College educated vs. Other.
- By baseline support: Three separate subgroups of bottom third, middle third, and top third of baseline support factor used to block

Canvasser Immigration Status

Below are results by the immigration status of the *canvasser*. All canvassers completed a demographic survey in which they were asked "Do you consider yourself to be an immigrant?" Responses are coded as 1 for yes and 0 for all other responses.

Table OA12: Heterogeneous treatment effects by Canvasser being an immigrant

	Canvasser being an immigrant = 1				Canvasser being an immigrant = 0			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.12	0.05	2.20	0.03	0.08	0.03	3.12	0.00
Abbrev	0.00	0.05	-0.09	0.93	0.03	0.03	1.15	0.25

Voter Closeness to Unauthorized Immigrants

Below are the results by the closeness of the respondent to unauthorized immigrants. In the baseline survey, we asked "Do you have any close friends, colleagues, or family members who are illegal or undocumented immigrants?". Responses are coded as 1 for yes and 0 for all other responses.

Table OA13: Heterogeneous treatment effects by Voter knows an unauthorized immigrant

	Voter knows an unauthorized immigrant = 1				Voter knows an unauthorized immigrant = 0			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.08	0.06	1.46	0.15	0.08	0.03	3.23	0.00
Abbrev	0.02	0.06	0.36	0.72	0.00	0.02	-0.06	0.95

Voter Immigration Status

Below are the results by the immigration status of the respondent. In the baseline survey, we asked "Which of these statements best describes you?" The provided responses were "I was born in the United States" and "I was born somewhere else". Responses are coded as 1 for being born in the US and 0 for all other responses.

Note that among the respondents to the one week survey who were born outside of the US, there were only 29 people in the full implementation condition, 30 in the abbreviated intervention condition, and 31 in the placebo condition, hence the noisy results.

Table OA14: Heterogeneous treatment effects by Voter born in the US

	Voter born in the US = 1				Voter born in the US = 0			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.09	0.02	3.83	0.00	-0.01	0.10	-0.14	0.89
Abbrev	0.01	0.02	0.40	0.69	0.16	0.09	1.78	0.08

By site

Below are the results by the canvass site. We noted in our pre-analysis plan that this was an exploratory analysis.

Table OA15: Heterogeneous treatment effects by site

	Orange County				Fresno				TN			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.04	0.04	0.92	0.36	0.12	0.05	2.58	0.01	0.13	0.04	3.50	0.0
Abbrev	-0.08	0.04	-1.91	0.06	0.10	0.05	2.24	0.02	0.02	0.04	0.52	0.6

By canvasser race

Below are the results by the race of the canvasser. We compare self-identified Latino canvassers to all others. We noted in our pre-analysis plan that this was an exploratory analysis.

Table OA16: Heterogeneous treatment effects by Latino canvasser

	Latino canvasser = 1				Latino canvasser = 0			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.07	0.03	2.14	0.03	0.10	0.03	2.98	0.0
Abbrev	-0.02	0.03	-0.70	0.48	0.06	0.03	1.63	0.1

By canvasser age

Below are the results by the age of the canvasser. We compare canvassers 30 and under vs. over 30. We noted in our pre-analysis plan that this was an exploratory analysis.

Table OA17: Heterogeneous treatment effects by Canvasser under 30

	Canvasser under 30 = 1				Canvasser under 30 = 0			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.07	0.03	2.58	0.01	0.20	0.05	3.77	0.00
Abbrev	-0.01	0.03	-0.34	0.74	0.12	0.05	2.54	0.01

By voter party

Below are the results by the party of the voter. We compare self-identified Democrats to Republicans to Independents (including leaners), as based on responses to the baseline survey. We noted in our pre-analysis plan that this was an exploratory analysis.

Table OA18: Heterogeneous treatment effects by voter party in baseline survey

	Democrat				Republican				Indep/Other			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.03	0.04	0.95	0.34	0.14	0.05	2.82	0.00	0.14	0.04	3.28	0.00
Abbrev	0.01	0.04	0.15	0.88	-0.03	0.04	-0.59	0.55	0.08	0.04	1.97	0.05

By voter race

Below are the results by the race of the voter. We compare self-identified Asian to Latino to White voters, as based on responses to the baseline survey. We noted in our pre-analysis plan that this was an exploratory analysis.

Note that among the respondents to the one week survey who were African American, there were only 16 people in the full implementation condition, 12 in the abbreviated intervention condition, and 13 in the placebo condition. Due to the small sample size, we do not include African American in the below table.

Table OA19: Heterogeneous treatment effects by voter race/ethnicity

	Asian				Latino				White			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.17	0.11	1.54	0.13	0.22	0.10	2.22	0.03	0.09	0.03	3.51	0.00
Abbrev	-0.07	0.13	-0.50	0.62	0.18	0.09	1.86	0.06	0.01	0.03	0.22	0.83

By voter economic status

Below are the results by the economic status of the voter. In the baseline survey, we asked "How would you rate your own personal financial situation?" Response options were "Excellent", "Good", "Only fair", "Poor", and "Would rather not say". We compare voters who said "Excellent" or "Good" to all other responses. We noted in our pre-analysis plan that this was an exploratory analysis.

Table OA20: Heterogeneous treatment effects by high financial status

	high financial status = 1				high financial status = 0			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.10	0.03	3.32	0.00	0.08	0.04	2.06	0.04
Abbrev	0.04	0.03	1.40	0.16	-0.05	0.04	-1.31	0.19

By voter education

Below are the results by the education status of the voter. We compare voters who self-reported having at least a college degree to all other voters. We noted in our pre-analysis plan that this was an exploratory analysis.

Table OA21: Heterogeneous treatment effects by College educated

	College educated = 1				College educated = 0			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.13	0.03	4.50	0.00	0.04	0.04	0.86	0.39
Abbrev	0.05	0.03	1.92	0.06	-0.04	0.04	-0.96	0.34

By voter baseline support

Below are the results by the baseline support of the voter. This is based on an index combining all of the unauthorized immigration policy and prejudice questions from the baseline survey. We divide this index into terciles and report results for each tercile. We noted in our pre-analysis plan that this was an exploratory analysis.

Table OA22: Heterogeneous treatment effects by immigration support in baseline survey

	Least Supportive				Mid Supportive				Most Suppotive			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Full	0.09	0.05	1.82	0.07	0.09	0.05	1.75	0.08	0.09	0.03	3.33	0.00
Abbrev	-0.04	0.05	-0.77	0.44	0.04	0.05	0.86	0.39	0.06	0.03	1.88	0.06

Estimates for Dichotomized Policy Items

It may be difficult for readers to interpret the magnitude of an effect presented in terms of standard deviation change. We therefore take two, non-pre-registered approaches to help communicate the substantive size of our estimates.

Strong Support

First, one way to make the results more interpretable is to examine treatment effects on whether participants said they strongly supported the policies asked about in the surveys. This attempts to recreate how participants might vote on each proposal if faced with a ballot measure or was deciding between candidates who differ on their immigration proposals. Note that we did not pre-specify this benchmarking procedure. We use this to illustrate the magnitude of our findings.

We first calculate the share of policies in the follow-up survey that individuals said they strongly supported. To calculate this, we dichotomize five of the policy items (all but `t#_imm_compassion` because this is a general as opposed to specific policy, although the effects on this item are larger), recoding them as 1 if the participant takes the most supportive immigration position in the first follow-up survey (one week post-treatment) and 0 for all other positions in the first follow-up survey. We then take the sum of these positions and compare across the conditions. Note that all estimated effects are intent-to-treat effects, and are not adjusted by the share of individuals who were contacted that actually had some or all of the conversations.

The average share of policies strongly supported in the Placebo group is 0.29.

Comparing the treatment conditions to the placebo, we find that the effect on the share of policies strongly supported of the Abbreviated Intervention condition is -0.005 with a p-value of 0.752; the effect on the share of policies strongly supported of the Full Intervention condition is 0.036 with a p-value of 0.011. The difference between the Full and Abbreviated condition is 0.038 with a p-value of 0.002. These statistics are covariate-adjusted using the same covariates and estimation approach as we pre-specified for the main analysis; although note again that this approach to dichotomizing the items was not pre-specified.

We also report results on the individual dichotimized items, set to 1 if an individual took a strong position on the supportive side of the issue and 0 otherwise.

The results on the individual dichotomized items are as follows:

Table OA23: At max at t1 (ITTs)

	With Covariates			
	Effect	SE	t.stat	p
Attorney				
Full vs. Placebo	0.041	0.021	1.906	0.057
Abbrev. vs. Placebo	0.014	0.022	0.651	0.515
Full vs. Abbrev.	0.024	0.023	1.076	0.282
DACA				
Full vs. Placebo	0.047	0.023	2.063	0.039
Abbrev. vs. Placebo	0.006	0.024	0.238	0.812
Full vs. Abbrev.	0.043	0.023	1.916	0.056
Deport All				
Full vs. Placebo	0.059	0.023	2.614	0.009
Abbrev. vs. Placebo	0.001	0.022	0.031	0.975
Full vs. Abbrev.	0.059	0.023	2.595	0.010
Citizenship				
Full vs. Placebo	0.027	0.024	1.135	0.257
Abbrev. vs. Placebo	-0.029	0.024	-1.230	0.219
Full vs. Abbrev.	0.055	0.024	2.304	0.021
Police				
Full vs. Placebo	0.007	0.022	0.298	0.766
Abbrev. vs. Placebo	-0.014	0.022	-0.618	0.537
Full vs. Abbrev.	0.021	0.023	0.910	0.363
Show Compassion				
Full vs. Placebo	0.061	0.022	2.789	0.005
Abbrev. vs. Placebo	0.024	0.022	1.100	0.271
Full vs. Abbrev.	0.037	0.022	1.648	0.100

Any Support

Second, we also conduct a version of this benchmarking where we dichotomize each variable to record whether participants registered any support (not only strong agreement). These new dichotomized variables are coded to 1 if a participant agreed at all with the policy and 0 otherwise (indicating stated indifference or opposition). In this analysis we again exclude the compassion item (where the effects are the largest but the policy is also not very specific).

The average share of policies supported at all in the Placebo group is 0.55. Comparing the treatment conditions to the placebo, we find that the effect on the share of policies supported at all of the Abbreviated Intervention condition is -0.012 with a p-value of 0.343; the effect on the share of policies strongly supported of the Full Intervention condition is 0.022 with a p-value of 0.058. The difference between the Full and Abbreviated conditions is 0.028 with a p-value of 0.007. These statistics are covariate-adjusted using the same covariates and estimation approach as we pre-specified for the main analysis; although note again that this approach to dichotomizing the items was not pre-specified.

The results on the individual items dichotimized in this manner are as follows:

Table OA24: Agree at all at t1 (ITTs)

	With Covariates			
	Effect	SE	t.stat	p
Attorney				
Full vs. Placebo	0.036	0.023	1.620	0.105
Abbrev. vs. Placebo	0.008	0.023	0.350	0.726
Full vs. Abbrev.	0.028	0.023	1.251	0.211
DACA				
Full vs. Placebo	0.029	0.021	1.392	0.164
Abbrev. vs. Placebo	-0.005	0.022	-0.218	0.827
Full vs. Abbrev.	0.034	0.021	1.635	0.102
Deport All				
Full vs. Placebo	0.027	0.022	1.247	0.213
Abbrev. vs. Placebo	-0.015	0.023	-0.683	0.495
Full vs. Abbrev.	0.044	0.022	2.022	0.043
Citizenship				
Full vs. Placebo	0.015	0.021	0.683	0.495
Abbrev. vs. Placebo	-0.032	0.023	-1.418	0.156
Full vs. Abbrev.	0.045	0.022	1.985	0.047
Police				
Full vs. Placebo	0.002	0.021	0.120	0.904
Abbrev. vs. Placebo	-0.013	0.022	-0.606	0.545
Full vs. Abbrev.	0.017	0.022	0.800	0.424
Show Compassion				
Full vs. Placebo	0.057	0.021	2.740	0.006
Abbrev. vs. Placebo	-0.003	0.022	-0.160	0.873
Full vs. Abbrev.	0.061	0.021	2.906	0.004

Results with Weights

To assess the generalizability of our results, we compare our main results – a sample average treatment effect (SATE) – to an estimate of the population average treatment effect (PATE). As [11] note, "The PATE can only be different from the SATE when two things hold: (1) there is meaningful variation in the treatment impact, and (2) that variation is correlated with the weights…It is important to compare the PATE and SATE estimates. A meaningful discrepancy between them is a signal to look for treatment effect heterogeneity and a flag that weight misspecification could be a real concern. If the estimates do not differ, however, and there is no other evidence of heterogeneity, then extrapolation is less of a concern – and furthermore the SATE is probably a sufficient estimate for the PATE."

To estimate the PATE, we first construct weights of who was canvassed and took the survey relative to the starting universe. We construct these weights using entropy balancing [8] and weight on gender, age, race, and vote history.

Below are results with and without these weights, showing that the estimated SATEs and PATEs are similar. If anything, the estimated PATEs are larger than the estimated SATEs, suggesting that the set of individuals who are canvassed and respond to surveys are perhaps more difficult to persuade than the broader universe.

Note that this analysis was not pre-registered but was prompted by feedback on the draft version of the paper.

Table OA25: ATE Effects on Overall Index with Weights

	Unweighted				Weighted			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Full vs. Placebo	0.089	0.024	3.792	0.000	0.112	0.032	3.486	0.001
Abbrev. vs. Placebo	0.016	0.023	0.664	0.507	0.022	0.029	0.733	0.464
1 Month								
Full vs. Placebo	0.068	0.024	2.774	0.006	0.100	0.032	3.130	0.002
Abbrev. vs. Placebo	0.035	0.024	1.487	0.137	0.030	0.027	1.092	0.275
3-6 Months								
Full vs. Placebo	0.070	0.027	2.612	0.009	0.100	0.036	2.816	0.005
Abbrev. vs. Placebo	0.018	0.026	0.696	0.487	0.023	0.032	0.728	0.467

References

[1] Emory Stephen Bogardus. A social distance scale. *Sociology & Social Research*, 1933.

[2] David Broockman and Joshua Kalla. Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282):220–224, 2016.

[3] David E Broockman, Joshua L Kalla, and Jasjeet S Sekhon. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis*, 25(4):435–464, 2017.

[4] Mark H Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113, 1983.

[5] Hunter Gehlbach and Maureen E. Brinkworth. The social perspective taking process: Strategies and sources of evidence in taking another's perspective. *Teachers College Record*, 114(1):1–29, 2012.

[6] Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation*. WW Norton, 2012.

[7] Alan S Gerber, Donald P Green, Edward H Kaplan, and Holger L Kern. Baseline, placebo, and treatment: Efficient estimation for three-group experiments. *Political Analysis*, 18(3):297–315, 2010.

[8] Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46, 2012.

[9] Seth J Hill and Gregory A Huber. On the meaning of survey reports of roll call votes not cast in a legislature. *American Journal of Political Science*, 2018.

[10] Gillian Ku, Cynthia S Wang, and Adam D Galinsky. The promise and perversity of perspective-taking in organizations. *Research in Organizational Behavior*, 35:79–102, 2015.

[11] Luke W. Miratrix, Jasjeet S. Sekhon, Alexander G. Theodoridis, and Luis F. Campos. Worth weighting? how to think about and use weights in survey experiments. *Political Analysis*, 26(3):275–291, 2018.

[12] David W Nickerson. Scalable protocols offer efficient design for field experiments. *Political Analysis*, 13(3):233–252, 2005.

[13] Richard E Petty and John T Cacioppo. *Communication and persuasion: Central and peripheral routes to attitude change*. Springer Series in Social Psychology, 1986.

Online Appendix for Experiments 2 and 3 (Transphobia)

Joshua Kalla[*]
David Broockman[†]

Contents

The full replication code and data that produces this report will be available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8BFYQO. This experiment was preregistered at Evidence in Governance and Politics (EGAP), see https://egap.org/registration/2005.

[*]Yale University, Departments of Political Science and Statistics & Data Science, josh.kalla@yale.edu
[†]University of California, Berkeley, Department of Political Science, dbroockman@berkeley.edu

Scripts

Experiment 2 (Canvass Experiment)

PAGE 1

PLACEBO SCRIPT - Plastic bags
Hi, is this _______ (THEIR NAME)? **[IF NO]** May I speak to ________? **[GATE KEEPER]** I'm a volunteer. May I please speak to __ because __ is on my list.
[ONLY IF PERSON ON LIST] Hi! I'm _____ (YOUR NAME), a volunteer calling from _________(CITY) to talk to registered voters about something we could vote on. How are you doing today? Great!
I'm volunteering with _________ (ORG) to talk to you about banning the practice of retailers giving plastic bags for free. If this law passed, retailers would have to charge customers 5 cents per bag. Is this something you support or oppose? (do not record answer) Thank you for sharing your opinion and have a great day!

LONG FORM SCRIPT - Nondiscrimination
Step 1: Initial Rating -- READ SLOWLY
Hi, is this _______ (THEIR NAME)? **[IF NO]** May I speak to ________? **[GATE KEEPER]** I'm a volunteer. May I please speak to __ because __ is on my list. **[ONLY IF PERSON ON LIST]** Hi! I'm _____ (YOUR NAME), a volunteer calling from _________(CITY) to talk to registered voters about something we could vote on. How are you doing today? Great!
I'm volunteering with _________ (ORG) to talk to you about ____'s (state) non-discrimination law. Under current state law it is still legal to fire an employee, evict someone, or deny them service in a restaurant, simply because the person is gay or transgender. Would you *support* or *oppose* updating our non-discrimination law to also protect gay and transgender people from unfair treatment in employment, housing, and places like hotels and restaurants? *****Okay, on a scale of 0-10, where 0 is totally sure you'd vote *against* updating the law and 10 is completely sure you'd vote in *favor*, where would you put yourself?** Why is that the right number for you? (listen)

Step 2: State your connection to transgender people
I'm volunteering today because: [pick one] • I am transgender. [share short coming out statement] --*OR*-- • I am gay/lesbian/bisexual and my good friend __________ is transgender. --*OR*-- • I have many gay and transgender friends and family members (like my friend ____________, who is transgender.)

Step 3: Connect with voter about about transgender people	
Do you personally know someone who is transgender?	
[If confusion on transgender]: I want to make sure we are on the same page about what we are talking about. A transgender person is someone who grows up knowing that their body doesn't match who they know they are on the inside, and so they transition and live every day as the gender they have always known themselves to be. For example, my friend _______ ….	
[IF NO - doesn't know a transgender person] Dig deeper to uncover concerns or questions.	**[IF YES - knows a transgender person] Dig deeper for details to develop empathy**
It can be hard to understand what it means to be transgender, especially if you've never met a transgender person. *--Based on what you've heard before, what do you think it means to be transgender?* *--Do you have any questions about it?*	How do you know them? What is their name? *--What was it like for you to learn that they are transgender? How did you feel?*

Version: 2016.08.05

[Show OPPO video to EVERY voter]

Now that you have seen that video, where 0 is totally *against* including gay and transgender people in non-discrimination laws and 10 is completely sure you'd vote in *favor*, **where would you put yourself?**

0	1	2	3	4	5	6	7	8	9	10
Against				Undecided			In Favor			

Why is that the right number for you?

6: DIG DEEPER: Process the ad, uncover concerns

This ad is scary, right?

[Triggered/Moved Down]	[Solid supporter]
--What concerns does it raise for you? *--What's on either side of this for you now?* *--Would you feel comfortable using/having a female member of your family use a restroom with a transgender woman, that is, someone who was born with a boy's body now lives their life every day as a woman?* *--What do you think this video is trying to say about transgender people?*	*--Why do you think that they made an ad talking about sexual predators?* *--What do you think they are trying to say about transgender people?* *--Is there any part of you that feels this ad could have a kernel of truth?*

7: ADDRESS CONCERNS: READ

So, I understand where you are coming from. [I think we are on the same page.] We all care about safety in restrooms. I have (*kids, grandkids, nieces*) and I would never support a law that I thought put them at risk.

But when we stop to think about this for a second, it's already illegal to harm someone, period. **And updating our nondiscrimination laws won't change that. I think we can protect people from discrimination and continue to hold offenders accountable.** *What do you think about that?*

8: PRO video: show and ask for reaction

Thank you for sharing your thoughts on that video. Here's a short one-minute video showing the perspective of a transgender person: **[Show PRO video to EVERY voter]**

What does this video make you think about?

- *How do you think it felt for the transgender woman in the video when the manager was trying to force her to use the men's room?*
- *How do you feel about the transgender woman's story? Thinking back to the video, what bathroom do you think she should be using? Why?*

What did you think about the two women who stepped in at the end to help the transgender woman? How did that make you feel?

9: Final Rating and Wrap Up

Now that we've been talking about this, if you were to vote tomorrow on a measure to update our nondiscrimination laws to include gay and transgender people, would you vote in favor or against?

To finish up, where 0 is totally *against* including gay and transgender people in non-discrimination laws and 10 is completely sure you'd vote in *favor*, **where would you put yourself?**

0	1	2	3	4	5	6	7	8	9	10
Against				Undecided			In Favor			

[IF MOVED] Why is that the right number? What makes you rate yourself differently?
[IF NOT MOVED] Why is that still the right number?

Thank you so much for your time. What is the best number to reach you at?

Experiment 3 (Phone Experiment)

PAGE 1

PLACEBO SCRIPT - Plastic bags
Hi, is this _______ (THEIR NAME)? **[IF NO]** May I speak to ________? **[GATE KEEPER]** I'm a volunteer. May I please speak to __ because __ is on my list. **[ONLY IF PERSON ON LIST]** Hi! I'm _____ (YOUR NAME), a volunteer calling from _________(CITY) to talk to registered voters about something we could vote on. How are you doing today? Great! I'm volunteering with _________ (ORG) to talk to you about banning the practice of retailers giving plastic bags for free. If this law passed, retailers would have to charge customers 5 cents per bag. Is this something you support or oppose? (do not record answer) Thank you for sharing your opinion and have a great day!

LONG FORM SCRIPT - Nondiscrimination
Step 1: Initial Rating -- READ SLOWLY
Hi, is this _______ (THEIR NAME)? **[IF NO]** May I speak to ________? **[GATE KEEPER]** I'm a volunteer. May I please speak to __ because __ is on my list. **[ONLY IF PERSON ON LIST]** Hi! I'm _____ (YOUR NAME), a volunteer calling from _________(CITY) to talk to registered voters about something we could vote on. How are you doing today? Great! I'm volunteering with _________ (ORG) to talk to you about ____'s (state) non-discrimination law. Under current state law it is still legal to fire an employee, evict someone, or deny them service in a restaurant, simply because the person is gay or transgender. Would you *support* or *oppose* updating our non-discrimination law to also protect gay and transgender people from unfair treatment in employment, housing, and places like hotels and restaurants? *****Okay, on a scale of 0-10, where 0 is totally sure you'd vote *against* updating the law and 10 is completely sure you'd vote in *favor*, where would you put yourself?** Why is that the right number for you? (listen)

Step 2: State your connection to transgender people
I'm volunteering today because: [pick one] ● I am transgender. [share short coming out statement] --*OR*-- ● I am gay/lesbian/bisexual and my good friend __________ is transgender. --*OR*-- ● I have many gay and transgender friends and family members (like my friend ____________, who is transgender.)

Step 3: Connect with voter about about transgender people	
Do you personally know someone who is transgender?	
[If confusion on transgender]: I want to make sure we are on the same page about what we are talking about. A transgender person is someone who grows up knowing that their body doesn't match who they know they are on the inside, and so they transition and live every day as the gender they have always known themselves to be. For example, my friend _______ ….	
[IF NO - doesn't know a transgender person] Dig deeper to uncover concerns or questions.	**[IF YES - knows a transgender person] Dig deeper for details to develop empathy**
It can be hard to understand what it means to be transgender, especially if you've never met a transgender person. *--Based on what you've heard before, what do you think it means to be transgender?* *--Do you have any questions about it?*	How do you know them? What is their name? *--What was it like for you to learn that they are transgender? How did you feel?*

Version: 2016.08.05

--How did you first learn about transgender people? What did you think then? How do you feel now? *--Do you know anyone gay/lesbian? How do you know them? Are you close to them? What is their name? How do you feel about them being gay/lesbian?*	*--What has their journey been like?* *--How are they treated?* *--How have your feelings changed?*

Step 4: Voter/You Share Judgement Stories

Updating our nondiscrimination law would protect gay and transgender people from unfair treatment. To me, this law is important because it gives us a chance to think about how we want to treat people who may seem different from us, in this case, gay and transgender people. What do you think about that?

Have you had an experience where someone treated you differently because of who you are or what other people thought about you? [If no] Have you witnessed this kind of judgment? When?

- *What was that like for you?*
- *What happened? Is there a specific time you remember being singled out?*
- *How did that feel?*
- *Did you tell anyone?*
- *Did anyone help you?*

[SHARE YOUR JUDGMENT STORY] *Thanks for sharing that, ____. I don't know what it is like to be in your shoes, but I can relate to the pain that judgment causes...I felt this when... (be vulnerable - tell story re: race, economic, language)*

Step 5: Introduce Bathrooms and Locker Room Concerns and RE-RATE

Thanks again for talking to me about this.

[IF SUPPORTIVE/UNDECIDED] I'm really glad you're supportive of *(not fully opposed to)* updating our law. However...
[IF OPPOSED] It sounds like you are pretty opposed to updating our nondiscrimination law to also protect gay and transgender people.

...we know some people have strong concerns, especially when it comes to public accommodations such as bathrooms in restaurants and shopping malls, or locker rooms in gyms.

They say that updating our laws would allow, and I'm using their language here, a *man* to say he feels like a *woman* to enter a restroom or locker room to sexually assault women and girls. **What do you think about that?** (Listen, do not respond.)

Some state legislatures refuse to consider updating our law because they believe registered sex offenders will take advantage of these laws. They suggest that allowing transgender women to enter these facilities creates uncertainty and danger in a so where women should be safe. So...

With this in mind, on that same 0-10 scale as before, where 0 is totally sure you'd vote *against* updating the law and 10 is completely sure you'd vote in *favor*, where would you put yourself? (listen) Why is that the right number for you?

Step 6: Dig Deeper on Concerns - Listen and Respond. Keep a curious, non-judgemental tone.

[Solid supporter 9-10] read some or all	**[Triggered/Moved Down/Not Supportive: 0-8]** read some or all
• *Is there any part of you that feels like these claims have even a kernel of truth?* • *Why do you think those who are opposed to updating the law are bringing up sexual predators in relation to a law that is supposed to protect gay and transgender people from discrimination? What do you think they are saying about transgender*	• *What concerns, if any, would you have about a transgender woman using women's restroom or locker room?* • *What's on either side of this for you now?* • *My friend _________, who I told you about earlier, is a transgender woman/man, who lives every day as a woman/man. Everyone at their workplace knows her/him as _________. What restroom do you think ________*

people? ● *How comfortable would you feel using the restroom or a locker room with a transgender person? What about (other) women and girls in your family; how comfortable do you think they would feel?*	*should use at work?* ● *I'm not sure I understand the connection between sexual predators and transgender people like my friend ____ I told you about earlier* (or their transgender acquaintance)*. Could you walk me through your thoughts on that?* ● *How comfortable would you feel using the restroom or a locker room with a transgender person? What about (other) women and girls in your family?*

Step 7: Address Concerns

So, I hear where you are coming from. [*I think we are on the same page.*] We all care about safety in restrooms. I have (*daughters, nieces, granddaughters*) and I would never support a law that I thought put them at risk. Does that make sense?

When I stopped to think about this for a second, it's already illegal enter a restroom or locker room to harm someone, period. And updating our nondiscrimination laws won't change that. I know we can protect people from discrimination and continue to hold offenders accountable. How do you feel about that? (Listen and respond)

[If still triggered] Is it comforting to you to hear that these kinds of laws have been around for a long time and that more than 200 cities and 19 states have enacted them with no increase in public safety incidents in any of these places?

Step 8: Link Back to Judgement Stories

Earlier, you shared a story about a time you felt judged, ___________________. And I could relate to that because of my experience _______________. So you and I both know what it feels like to be treated unfairly.

To me, (as I said before) updating this law is important because it gives us a chance to think about how we want to treat one another.

(read some or all, as appropriate)

- *How do you think a transgender person would feel being told to leave a restroom that matches the gender they live every day?*
- *What about if they were fired from their job because their boss was uncomfortable with the fact that they're transgender, despite the fact that they work hard and do their job well?*
- *If you saw a boss or coworker making fun of or judging a gay or transgender person, what do you think you would do in that situation?* (listen and respond)

Step 9: RE-RATE and goodbye

Thank you so much for talking to me today. Now that we've been talking about this, if you were to vote tomorrow on whether or not to update our nondiscrimination laws to also protect gay and transgender people, would you vote in favor or against?

One last time, on that same 0-10 scale as before, where 0 is totally sure you'd vote *against* updating the law and 10 is completely sure you'd vote in *favor*, where would you put yourself?

- [If Moved] Why is that the right number? What makes you rate yourself differently?
- [If Not Moved] Why is that still the right number?

Thank you so much for your time. I enjoyed our conversation! Have a great night.

Survey Recruitment Procedures and Experimental Design

The general survey recruitment procedures and experimental design were identical to Experiment 1 except as otherwise noted below.

Experiments 2 and 3 had the following steps:

- We first attempted to survey 324615 voters in the target universes.
- 17252 voters responded to this survey from 14935 households.
- For Experiment 2, at the household level, these respondents were randomly assigned to receive the Participants' and Video Narratives Condition (1095), the Video Narratives Only Condition (1214), or to a placebo condition about recycling (1176).
- The implementing partners canvassed 384 voters with the Participants' and Video Narratives Condition, 457 voters with the Video Narratives Only Condition, and 493 with a placebo conversation.
- For Experiment 3, we took individuals who either (a) lived outside of the canvass area or (b) were never reached during the canvass phase (e.g., a canvasser never knocked on the door or nobody was home) and randomly assigned them to receive either a phone conversation with the Participants' Narratives Intervention (6892) or a placebo phone conversation about recycling (6898).
- The implementing partners spoke with 1268 voters in the phone intervention and 1369 with placebo calls.
- We then successfully resurveyed 75% of the voters they contacted successfully in a survey one week following the intervention and 73% of the voters they contacted successfully in a survey one month following the intervention.

The below table gives the raw counts for each conditions and other ancillary statistics:

Metric	Canvass Placebo	Canvass Participants' and Video Narratives Condition	Canvass Video Narratives Only	Phone Placebo	Phone Participants' Narratives
Number Assigned	2823	2815	2815	6898	6892
Results: Number Reached	653	556	649	1369	1268
Reached 1st Rating	n/a	294	349	n/a	766
Reached 2nd Rating	n/a	239	304	n/a	527
Reached 3rd Rating	n/a	128	146	n/a	489
Length: 0-2 minutes	208	41	49	1304	454
Length: 2-5 minutes	215	48	88	50	218
Length: 5-10 minutes	17	77	159	9	283
Length: 10-15 minutes	5	92	87	4	189

Metric	Canvass Placebo	Canvass Participants' and Video Narratives Condition	Canvass Video Narratives Only	Phone Placebo	Phone Participants' Narratives
Length: Over 15 minutes	0	60	34	2	70
Voter Shared Story	n/a	202	44	n/a	368
Canvasser Shared Story	n/a	251	69	n/a	477

Outcomes

We conducted two follow-up surveys for Experiments 2 and 3: a first one week after contact and a second one month after contact. We asked voters 15 questions on each survey, 9 about their views towards transgender-inclusive non-discrimination laws and 6 about their tolerance of transgender people.

Working together with the implementation partners, we developed the following outcome measures that appeared on the survey. The survey items were split into two categories: prejudice items and policy items.

During the surveys, we also showed an opposition political ad closely patterned after an ad run by opponents of transgender-inclusive non-discrimination laws in Houston in 2015 alleging that sexual predators would abuse transgender-inclusive non-discrimination laws. This video was included to measure whether treatment effects persist after exposure to counter-arguments [4]. Half of our outcome measures were asked before this video was shown and half were asked after.

Prejudice Items Before Video

- `transprej_comfortwork` I would feel comfortable working closely with a transgender person (a person who was born with a boy's body but now identifies as a woman or a person who was born with a girl's body but now identifies as a man).
- `transprej_moralwrong` Saying you are a gender that is different than the one you were born as is morally wrong.
- `transprej_moralgenderchange` It is morally wrong for someone born with a boy's body to undergo a gender change and live every day as a woman.
- `transprej_friend` I would support a friend choosing to have a sex change.
- `transprej_restroom` It would be wrong to allow a transgender woman (a person who was born with a boy's body but identifies as a woman) to use a woman's restroom or locker room.

Policy Items Before Video

- `transpolicy_LGBTdiscrim` A law in your state that would protect gay and transgender people from discrimination in employment, housing, and public accommodations.
- `transpolicy_lawbathroom` (Starting on second survey.) Our state's nondiscrimination law should allow transgender people to use the restroom that matches the gender they live every day—so a person who lives every day as a woman could use the women's restroom, even if that person was born and raised as a boy.

Prejudice Items After Video

- `transvid_comfortbathroom` I would feel comfortable sharing a bathroom with someone who is transgender.
- `therm_trans` Rating of transgender people on a feeling thermometer.
- `transvid_restroom` It would be wrong to allow a transgender woman (a person who was born with a boy's body but identifies as a woman) to use a woman's restroom or locker room.

Policy Items After Video

- `transvid_LGBTdiscrim` A law in your state that would protect gay and transgender people from discrimination in employment, housing, and public accommodations.
- `transvid_fire` A law protecting transgender people from being fired for being transgender.
- `transvid_school` A law that requires transgender students to use the school bathrooms and locker rooms that match their biological or anatomical sex at birth, rather than the gender they live as every day.
- `transvid_predator` I'm concerned that sexual predators could take advantage of a nondiscrimination law to put women's and children's safety at risk.
- `transvid_teacher` Transgender people should not be allowed to serve as public school teachers.

Computing Indices

We computed our indices using the below code.

```
compute.index.dv <- function(dv.names, survey.wave.boolean.vector){
  responders <- analysis.data[survey.wave.boolean.vector==1,]
  dv.names <- paste0(dv.names, '_scaled')
  index <- rowMeans(responders[,dv.names], na.rm = TRUE)
  index <- scale(index)
  return(index[match(analysis.data$id, responders$id)])
}
```

Estimation Procedures

Consistent with our pre-analysis plan, to estimate treatment effects we use ordinary least squares (OLS) regressions with cluster-robust standard errors, clustering on household and also including the pre-treatment covariates from the baseline survey and voter list named in our pre-analysis plan. This procedure and these covariates were pre-specified in advance and produce unbiased estimates of causal effects [2, 1]. Note that there is no reclassification of treatment based on what occurs at the door and we do not exclude any subjects who came to the door; we compare all subjects who came to the door and were pre-assigned to the treatment conversation to all subjects who came to the door and were pre-assigned to the placebo conversation.

Tests of Design Assumptions

Survey Representativeness

The below tables shows how the representativeness of those who responded to the survey differ from those mailed an invitation to participate in the survey. These data come from the voter file. Note that no weighting is used in the analysis; the aim of the estimation is to test for the existence of treatment effects within this sample, not to generalize to the population of invited respondents.

This first table examines the canvass experiment (Experiment 2).

Table OA26: Representativeness of Canvass Experiment at Each Stage

Sample	Female	Reg. Dem	Reg. Rep	Af-Am	Latino	White	Voted 14	Voted 12	Voted 10	Voted 08	AZ	FL	GA	OH	N
Starting	0.55	0.27	0.3	0.1	0.02	0.82	0.63	0.87	0.64	0.83	0.26	0.25	0.22	0.27	159941
Baseline Resp.	0.54	0.3	0.27	0.05	0.02	0.87	0.77	0.91	0.73	0.85	0.33	0.21	0.26	0.2	8440
Canvassed	0.53	0.34	0.28	0.06	0.02	0.88	0.83	0.93	0.79	0.88	0.27	0.25	0.25	0.23	1858
1 Wk Resp.	0.52	0.33	0.27	0.05	0.02	0.89	0.82	0.93	0.77	0.86	0.25	0.29	0.25	0.21	1044
1 Mo Resp.	0.53	0.34	0.26	0.05	0.02	0.88	0.83	0.92	0.77	0.86	0.25	0.29	0.25	0.21	989

This second table examines the phone experiment (Experiment 3).

Table OA27: Representativeness of Phone Experiment at Each Stage

Sample	Female	Reg. Dem	Reg. Rep	Af-Am	Latino	White	Voted 14	Voted 12	Voted 10	Voted 08	AZ	FL	GA	OH	N
Starting	0.54	0.24	0.36	0.05	0.02	0.88	0.76	0.94	0.78	0.91	0.27	0.22	0.25	0.26	169638
Baseline Resp.	0.52	0.29	0.33	0.04	0.02	0.91	0.84	0.95	0.82	0.91	0.37	0.18	0.23	0.22	13767
Called	0.53	0.32	0.34	0.04	0.01	0.92	0.91	0.97	0.9	0.95	0.32	0.16	0.22	0.3	2637
1 Wk Resp.	0.52	0.33	0.33	0.03	0.01	0.93	0.91	0.98	0.91	0.95	0.32	0.16	0.21	0.31	1943
1 Mo Resp.	0.53	0.33	0.34	0.03	0.01	0.93	0.92	0.98	0.91	0.96	0.32	0.16	0.21	0.31	1897

Covariate Balance among All Subjects, Compliers, and Reporters in Canvass Experiment

The below tables demonstrate that balance on pre-treatment observable attributes is maintained among the original universe of pre-survey respondents randomized to each group, the sub-sample that was canvassed, and the sub-sample that was both canvassed and successfully re-interviewed for the canvass experiment. Each table shows the mean value for the covariate under each condition as well as the *p*-value from a one-way ANOVA test. The first table considers all voters who were randomly assigned after having taken the pre-survey (all subjects); the second table considers all voters who were successfully contacted (compliers); the remaining tables consider all voters who responded to the first and second post-surveys (reporters).

Table OA28: Covariate Balance among Pre-Survey Respondents, Canvass Experiment

	Placebo	Video Narratives Only Condition	Participants' and Video Narratives Condition	p-value
Registered Democrat	0.29	0.29	0.30	0.7
Registered Republican	0.28	0.27	0.27	0.9
Female	0.52	0.56	0.55	0.05
White	0.88	0.86	0.87	0.06
Transgender People Feeling Thermometer t0	58.35	58.51	58.18	0.89
Donald Trump Feeling Thermometer t0	28.40	28.90	27.81	0.47
Barack Obama Feeling Thermometer t0	54.28	54.41	54.19	0.98
Hillary Clinton Feeling Thermometer t0	42.55	42.99	42.30	0.77
N	2818.00	2811.00	2811.00	-

Table OA29: Covariate Balance among Compliers, Canvass Experiment

	Placebo	Video Narratives Only Condition	Participants' and Video Narratives Condition	p-value
Registered Democrat	0.33	0.34	0.34	0.89
Registered Republican	0.28	0.27	0.28	0.96
Female	0.51	0.55	0.54	0.24
White	0.89	0.86	0.88	0.22
Transgender People Feeling Thermometer t0	58.09	58.15	57.54	0.91
Donald Trump Feeling Thermometer t0	28.01	29.75	29.22	0.64
Barack Obama Feeling Thermometer t0	54.33	54.01	53.22	0.88
Hillary Clinton Feeling Thermometer t0	44.09	44.43	42.56	0.65
N	653.00	649.00	556.00	-

Table OA30: Covariate Balance among 1st Post-Survey Respondents, Canvass Experiment

	Placebo	Video Narratives Only Condition	Participants' and Video Narratives Condition	p-value
Registered Democrat	0.32	0.35	0.32	0.67
Registered Republican	0.28	0.26	0.28	0.82
Female	0.52	0.53	0.53	0.92
White	0.90	0.87	0.90	0.28
Transgender People Feeling Thermometer t0	59.76	59.79	58.18	0.66
Donald Trump Feeling Thermometer t0	28.21	25.84	29.47	0.36
Barack Obama Feeling Thermometer t0	54.43	57.26	53.70	0.45
Hillary Clinton Feeling Thermometer t0	43.10	46.21	42.16	0.32
N	387.00	352.00	305.00	-

Table OA31: Covariate Balance among 2nd Post-Survey Respondents, Canvass Experiment

	Placebo	Video Narratives Only Condition	Participants' and Video Narratives Condition	p-value
Registered Democrat	0.34	0.34	0.36	0.86
Registered Republican	0.26	0.26	0.27	0.98
Female	0.52	0.55	0.52	0.69
White	0.90	0.87	0.88	0.43
Transgender People Feeling Thermometer t0	59.67	60.66	59.68	0.85
Donald Trump Feeling Thermometer t0	27.89	25.71	28.24	0.58
Barack Obama Feeling Thermometer t0	55.72	57.52	55.69	0.78
Hillary Clinton Feeling Thermometer t0	44.23	46.66	43.54	0.52
N	369.00	334.00	286.00	-

Covariate Balance among All Subjects, Compliers, and Reporters in Phone Experiment

The below tables demonstrate that balance on pre-treatment observable attributes is maintained among the original universe of pre-survey respondents randomized to each group, the sub-sample that was called, and the sub-sample that was both called and successfully re-interviewed for the phone experiment. Each table shows the mean value for the covariate under each condition as well as the *p*-value from a t-test, given there are only two groups. The first table considers all voters who were randomly assigned after having taken the pre-survey (all subjects); the second table considers all voters who were successfully contacted (compliers); the remaining tables consider all voters who responded to the first and second post-surveys (reporters).

Table OA32: Covariate Balance among Pre-Survey Respondents, Phone Experiment

	Placebo	Phone Intervention	p-value
Registered Democrat	0.28	0.29	0.22
Registered Republican	0.33	0.33	0.96
Female	0.52	0.52	0.71
White	0.90	0.91	0.72
Transgender People Feeling Thermometer t0	56.24	56.21	0.95
Donald Trump Feeling Thermometer t0	31.51	31.32	0.74
Barack Obama Feeling Thermometer t0	50.08	49.97	0.87
Hillary Clinton Feeling Thermometer t0	40.23	40.03	0.75
N	6888.00	6879.00	-

Survey Attrition in Canvass Experiment

An important design assumption is that the treatment does not affect the composition of the individuals who take each follow-up survey [1]. We investigate this by regressing an indicator for responding to a post-treatment survey on indicators of treatment assignment. Across the two survey waves, we find slight

Table OA33: Covariate Balance among Compliers, Phone Experiment

	Placebo	Phone Intervention	p-value
Registered Democrat	0.31	0.34	0.15
Registered Republican	0.34	0.33	0.51
Female	0.52	0.55	0.16
White	0.92	0.92	0.57
Transgender People Feeling Thermometer t0	54.38	56.53	0.03
Donald Trump Feeling Thermometer t0	33.98	32.38	0.25
Barack Obama Feeling Thermometer t0	48.05	50.96	0.06
Hillary Clinton Feeling Thermometer t0	40.04	41.92	0.2
N	1369.00	1268.00	-

Table OA34: Covariate Balance among 1st Post-Survey Respondents, Phone Experiment

	Placebo	Phone Intervention	p-value
Registered Democrat	0.31	0.34	0.14
Registered Republican	0.34	0.33	0.75
Female	0.51	0.53	0.38
White	0.94	0.93	0.49
Transgender People Feeling Thermometer t0	55.38	57.15	0.13
Donald Trump Feeling Thermometer t0	31.58	30.60	0.53
Barack Obama Feeling Thermometer t0	49.59	52.35	0.13
Hillary Clinton Feeling Thermometer t0	41.05	43.00	0.25
N	1014.00	929.00	-

Table OA35: Covariate Balance among 2nd Post-Survey Respondents, Phone Experiment

	Placebo	Phone Intervention	p-value
Registered Democrat	0.32	0.34	0.31
Registered Republican	0.34	0.34	0.77
Female	0.52	0.54	0.21
White	0.94	0.93	0.46
Transgender People Feeling Thermometer t0	55.02	57.35	0.05
Donald Trump Feeling Thermometer t0	31.84	30.42	0.37
Barack Obama Feeling Thermometer t0	49.43	52.66	0.08
Hillary Clinton Feeling Thermometer t0	40.48	43.13	0.12
N	994.00	903.00	-

evidence of differential attrition. In the below sections, we show that pre-treatment covariates do not predict this slight differential attrition.

Table OA36: Test for differential attrition, Canvass Experiment

	Effect	SE	t.stat	p
1 Week				
Video Narratives Only Condition	-0.05	0.03	-1.83	0.07
Participants' and Video Narratives Condition	-0.04	0.03	-1.54	0.12
1 Month				
Video Narratives Only Condition	-0.05	0.03	-1.82	0.07
Participants' and Video Narratives Condition	-0.05	0.03	-1.76	0.08

Survey Attrition in Phone Experiment

In the below table we look at the phone experiment and again, across the two survey waves, we find no evidence of differential attrition.

Table OA37: Test for differential attrition, Phone Experiment

	Effect	SE	t.stat	p
1 Week				
Treat	-0.01	0.02	-0.47	0.64
1 Month				
Treat	-0.01	0.02	-0.80	0.43

Test of Differential Attrition by Covariates in Canvass Experiment

The above subsection demonstrated that there was no average differential attrition; now, we test for whether the treatment caused attrition to differ by covariates (for example, whether it encouraged already-supportive subjects to complete the post-survey but also discouraged unsupportive subjects from doing so) [2]. To test whether attrition patterns are similar by covariates in treatment and placebo, we use a linear regression of whether or not an individual responded to the follow-up survey on treatment, baseline covariates, and treatment-covariate interactions. We then perform a heteroskedasticity-robust F-test of the hypothesis that all the interaction coefficients are zero. Below we report the p-value of this F-test. Based on the results presented below, there does not appear to be evidence of asymmetrical attrition.

Table OA38: p-value by Survey Wave Test of Differential Attrition by Covariates, Canvass Experiment

1 Week Survey (t1)	0.68
1 Month Survey (t2)	0.38

Test of Differential Attrition by Covariates in Phone Experiment

Below we present the same test for the phone experiment.

Table OA39: p-value by Survey Wave Test of Differential Attrition by Covariates, Phone Experiment

1 Week Survey (t1)	0.75
1 Month Survey (t2)	0.53

Canvass Results (Experiment 2)

Effects on Overall Index

First, we show the effects on an overall index that combines all the outcomes above together. Overall, we see statistically significant effects from all types of conversations and that these effects persist for at least one month.

Table OA40: ATE effects on overall index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' and Video Narratives Condition vs. Placebo	0.093	0.027	3.494	0.000	0.021	0.079	0.267	0.790
Video Narratives Only Condition vs. Placebo	0.094	0.024	3.969	0.000	0.047	0.075	0.619	0.536
Participants' and Video Narratives vs. Video Narratives Only	-0.002	0.028	-0.062	0.951	-0.025	0.082	-0.308	0.758
1 Month								
Participants' and Video Narratives Condition vs. Placebo	0.080	0.026	3.049	0.002	0.064	0.080	0.804	0.422
Video Narratives Only Condition vs. Placebo	0.064	0.025	2.513	0.012	0.041	0.077	0.531	0.596
Participants' and Video Narratives vs. Video Narratives Only	0.016	0.030	0.538	0.591	0.023	0.083	0.276	0.783
Pooled								
Participants' and Video Narratives Condition vs. Placebo	0.079	0.024	3.346	0.001	0.043	0.076	0.564	0.573
Video Narratives Only Condition vs. Placebo	0.079	0.022	3.572	0.000	0.031	0.073	0.426	0.670
Participants' and Video Narratives vs. Video Narratives Only	-0.001	0.026	-0.022	0.983	0.012	0.080	0.148	0.882

Effects on Policy Index

Next, we show the effects on the policy index.

Table OA41: ATE effects on policy index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' and Video Narratives Condition vs. Placebo	0.089	0.034	2.660	0.008	0.029	0.078	0.367	0.714
Video Narratives Only Condition vs. Placebo	0.072	0.031	2.359	0.018	0.036	0.074	0.488	0.626
Participants' and Video Narratives vs. Video Narratives Only	0.018	0.035	0.501	0.617	-0.007	0.082	-0.090	0.928
1 Month								
Participants' and Video Narratives Condition vs. Placebo	0.070	0.032	2.217	0.027	0.057	0.079	0.721	0.471
Video Narratives Only Condition vs. Placebo	0.054	0.030	1.807	0.071	0.028	0.077	0.361	0.718
Participants' and Video Narratives vs. Video Narratives Only	0.018	0.035	0.498	0.618	0.030	0.084	0.351	0.726
Pooled								
Participants' and Video Narratives Condition vs. Placebo	0.073	0.029	2.522	0.012	0.041	0.076	0.546	0.585
Video Narratives Only Condition vs. Placebo	0.065	0.028	2.369	0.018	0.020	0.072	0.282	0.778
Participants' and Video Narratives vs. Video Narratives Only	0.010	0.031	0.303	0.762	0.021	0.080	0.260	0.795

Effects on Prejudice Index

Next, we show the effects on the prejudice index.

Table OA42: ATE effects on prejudice index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' and Video Narratives Condition vs. Placebo	0.091	0.028	3.308	0.001	0.015	0.080	0.184	0.854
Video Narratives Only Condition vs. Placebo	0.104	0.025	4.213	0.000	0.052	0.076	0.683	0.495
Participants' and Video Narratives vs. Video Narratives Only	-0.015	0.029	-0.535	0.593	-0.037	0.082	-0.448	0.654
1 Month								
Participants' and Video Narratives Condition vs. Placebo	0.087	0.029	3.010	0.003	0.068	0.080	0.844	0.399
Video Narratives Only Condition vs. Placebo	0.070	0.027	2.629	0.009	0.053	0.078	0.687	0.492
Participants' and Video Narratives vs. Video Narratives Only	0.014	0.031	0.435	0.664	0.014	0.082	0.176	0.860
Pooled								
Participants' and Video Narratives Condition vs. Placebo	0.083	0.025	3.335	0.001	0.045	0.077	0.582	0.561
Video Narratives Only Condition vs. Placebo	0.089	0.023	3.925	0.000	0.041	0.074	0.558	0.577
Participants' and Video Narratives vs. Video Narratives Only	-0.008	0.027	-0.314	0.753	0.004	0.080	0.048	0.962

Effects on Pre-Video Index

As described above, halfway through the surveys we showed an opposition political ad closely patterned after an ad run by opponents of transgender-inclusive non-discrimination laws in Houston in 2015 alleging that sexual predators would abuse transgender-inclusive non-discrimination laws. This video was included to measure whether treatment effects persist after exposure to counter-arguments [4].

Next, we show the effects on the index of questions asked before the opposition video was shown. This is a mix of prejudice and policy questions.

Table OA43: ATE effects on pre-video index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' and Video Narratives Condition vs. Placebo	0.066	0.028	2.362	0.018	-0.019	0.080	-0.240	0.810
Video Narratives Only Condition vs. Placebo	0.074	0.027	2.758	0.006	0.015	0.075	0.198	0.843
Participants' and Video Narratives vs. Video Narratives Only	-0.009	0.030	-0.310	0.757	-0.034	0.082	-0.414	0.679
1 Month								
Participants' and Video Narratives Condition vs. Placebo	0.084	0.029	2.920	0.004	0.060	0.080	0.746	0.456
Video Narratives Only Condition vs. Placebo	0.072	0.028	2.627	0.009	0.041	0.077	0.527	0.598
Participants' and Video Narratives vs. Video Narratives Only	0.011	0.031	0.352	0.725	0.019	0.082	0.229	0.819
Pooled								
Participants' and Video Narratives Condition vs. Placebo	0.070	0.024	2.894	0.004	0.022	0.077	0.292	0.770
Video Narratives Only Condition vs. Placebo	0.075	0.023	3.198	0.001	0.018	0.073	0.249	0.804
Participants' and Video Narratives vs. Video Narratives Only	-0.006	0.026	-0.216	0.829	0.004	0.080	0.055	0.956

Effects on Post-Video Index

Next, we show the effects on the index of questions asked after the opposition video was shown. This is a mix of prejudice and policy questions.

Table OA44: ATE effects on post-video index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' and Video Narratives Condition vs. Placebo	0.110	0.031	3.529	0.000	0.052	0.078	0.663	0.507
Video Narratives Only Condition vs. Placebo	0.106	0.027	3.855	0.000	0.070	0.076	0.920	0.358
Participants' and Video Narratives vs. Video Narratives Only	0.004	0.033	0.129	0.897	-0.017	0.082	-0.212	0.832
1 Month								
Participants' and Video Narratives Condition vs. Placebo	0.075	0.030	2.490	0.013	0.067	0.080	0.835	0.404
Video Narratives Only Condition vs. Placebo	0.056	0.028	1.973	0.049	0.040	0.077	0.520	0.603
Participants' and Video Narratives vs. Video Narratives Only	0.019	0.034	0.548	0.584	0.026	0.084	0.314	0.753
Pooled								
Participants' and Video Narratives Condition vs. Placebo	0.083	0.027	3.062	0.002	0.058	0.076	0.762	0.446
Video Narratives Only Condition vs. Placebo	0.082	0.026	3.191	0.002	0.041	0.074	0.559	0.576
Participants' and Video Narratives vs. Video Narratives Only	0.003	0.030	0.094	0.925	0.017	0.080	0.208	0.835

Phone Results (Experiment 3)

Effects on Overall Index

First, we show the effects on an overall index that combines all the outcomes above together. Overall, we see statistically significant effects from all types of conversations and that these effects persist for at least one month.

Table OA45: ATE effects on overall index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' Narratives Condition vs. Placebo	0.046	0.015	3.115	0.002	0.080	0.046	1.738	0.082
1 Month								
Participants' Narratives Condition vs. Placebo	0.044	0.016	2.780	0.006	0.114	0.047	2.438	0.015
Pooled								
Participants' Narratives Condition vs. Placebo	0.045	0.014	3.247	0.001	0.090	0.044	2.038	0.042

Effects on Policy Index

Next, we show the effects on the policy index.

Table OA46: ATE effects on policy index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' Narratives Condition vs. Placebo	0.034	0.019	1.829	0.068	0.073	0.046	1.568	0.117
1 Month								
Participants' Narratives Condition vs. Placebo	0.029	0.018	1.592	0.112	0.098	0.047	2.098	0.036
Pooled								
Participants' Narratives Condition vs. Placebo	0.033	0.016	1.999	0.046	0.080	0.044	1.815	0.070

Effects on Prejudice Index

Next, we show the effects on the prejudice index.

Table OA47: ATE effects on prejudice index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' Narratives Condition vs. Placebo	0.052	0.016	3.196	0.001	0.081	0.046	1.770	0.077
1 Month								
Participants' Narratives Condition vs. Placebo	0.058	0.018	3.308	0.001	0.125	0.046	2.689	0.007
Pooled								
Participants' Narratives Condition vs. Placebo	0.054	0.015	3.601	0.000	0.095	0.044	2.166	0.030

Effects on Pre-Video Index

As described above, halfway through the surveys we showed an opposition political ad closely patterned after an ad run by opponents of transgender-inclusive non-discrimination laws in Houston in 2015 alleging that sexual predators would abuse transgender-inclusive non-discrimination laws. This video was included to measure whether treatment effects persist after exposure to counter-arguments [4].

Next, we show the effects on the index of questions asked before the opposition video was shown. This is a mix of prejudice and policy questions.

Table OA48: ATE effects on pre-video index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' Narratives Condition vs. Placebo	0.036	0.017	2.091	0.037	0.060	0.046	1.298	0.195
1 Month								
Participants' Narratives Condition vs. Placebo	0.050	0.018	2.800	0.005	0.111	0.047	2.379	0.017
Pooled								
Participants' Narratives Condition vs. Placebo	0.046	0.015	2.989	0.003	0.083	0.044	1.875	0.061

Effects on Post-Video Index

Next, we show the effects on the index of questions asked after the opposition video was shown. This is a mix of prejudice and policy questions.

Table OA49: ATE effects on post-video index

	With Covariates				Without Covariates			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' Narratives Condition vs. Placebo	0.052	0.017	3.076	0.002	0.092	0.046	2.006	0.045
1 Month								
Participants' Narratives Condition vs. Placebo	0.039	0.018	2.219	0.027	0.112	0.047	2.402	0.016
Pooled								
Participants' Narratives Condition vs. Placebo	0.043	0.015	2.824	0.005	0.093	0.044	2.110	0.035

Other Results

Results with Weights

As described in the SM for Experiment 1, to assess the generalizability of our results, we compare our main results – a sample average treatment effect (SATE) – to an estimate of the population average treatment effect (PATE). To estimate the PATE, we first construct weights of who was canvassed and took the survey relative to the starting universe. We construct these weights using entropy balancing [3] and weight on gender, age, race, party registration, and vote history.

Below are results with and without these weights, showing that the estimated SATEs and PATEs are similar. If anything, the estimated PATE is usually larger than the SATE, suggesting that the set of individuals who are canvassed and respond to surveys are perhaps more difficult to persuade than the broader universe.

Note that this analysis was not pre-registered but was prompted by feedback on the draft version of the paper.

Canvass

Table OA50: ATE Effects on Overall Index with Weights, Canvass Experiment

	Unweighted				Weighted			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' and Video Narratives Condition vs. Placebo	0.093	0.027	3.494	0.000	0.111	0.029	3.821	0.000
Video Narratives Only Condition vs. Placebo	0.094	0.024	3.969	0.000	0.093	0.029	3.201	0.001
1 Month								
Participants' and Video Narratives Condition vs. Placebo	0.080	0.026	3.049	0.002	0.109	0.028	3.859	0.000
Video Narratives Only Condition vs. Placebo	0.064	0.025	2.513	0.012	0.060	0.030	1.978	0.048

Phone

Table OA51: ATE Effects on Overall Index with Weights, Phone Experiment

	Unweighted				Weighted			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p
1 Week								
Participants' Narratives Condition vs. Placebo	0.046	0.015	3.115	0.002	0.038	0.018	2.148	0.032
1 Month								
Participants' Narratives Condition vs. Placebo	0.044	0.016	2.781	0.005	0.047	0.020	2.327	0.020

Subgroups

In our pre-analysis plan, we specified that we would examine treatment effect heterogeneity by:

- Whether the ATE of canvassing is different for canvassers who identify as transgender or gender non-conforming than for all other canvassers,
- Whether the ATE of canvassing is different for canvassers who are perceived as gender conforming,
- Whether the ATE of canvassing is different for paid vs. volunteer staff, and
- Whether the ATE of canvassing and calling varies by the political knowledge of the participant.

While we did not specify this in the pre-analysis plan, we will also investigate treatment effect heterogeneity by the party identification of the participant.

Note that we only collected the demographics of canvassers, not the callers. We collected the canvasser demographics via survey were they described their gender identity and how they anticipate others perceive their gender identity.

For this subgroup analysis, we present ATE results on the overall index in the 1 week survey.

By canvasser gender identity

In the canvass, 384 conversations and one week surveys were completed by canvassers who self-identify as transgender or gender non-conforming; 596 by canvassers who self-identify as cis-gender; and 64 by canvassers for whom we are missing data.

Table OA52: Heterogeneous treatment effects by gender identity

	Transgender or Gender Non-Conforming				Cisgender				Missing Data			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Participants' and Video Narratives	0.11	0.04	2.89	0.0	0.07	0.04	1.98	0.05	0.19	0.15	1.24	0.22
Video Narratives Only	0.07	0.04	1.63	0.1	0.09	0.03	3.07	0.00	0.24	0.11	2.24	0.03

By canvasser gender perception

In the canvass, 550 conversations and one week surveys were completed by canvassers who self-identify as being perceived as gender conforming; 430 by canvassers who self-identify as being perceived as gender non-conforming; and 64 by canvassers for whom we are missing data.

Table OA53: Heterogeneous treatment effects by gender identity perception

	Perceived as gender conforming				Perceived as gender non-conforming				Missing Data			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Participants' and Video Narratives	0.08	0.04	2.02	0.04	0.10	0.04	2.42	0.02	0.19	0.15	1.24	0.22
Video Narratives Only	0.06	0.03	1.87	0.06	0.11	0.04	2.93	0.00	0.24	0.11	2.24	0.03

By canvasser volunteer status

In the canvass, 422 conversations and one week surveys were completed by paid canvassers; 558 by volunteer canvassers; and 64 by canvassers for whom we are missing data.

Table OA54: Heterogeneous treatment effects by canvasser volunteer status

	Paid canvasser				Volunteer canvasser				Missing Data			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Participants' and Video Narratives	0.13	0.04	2.85	0.00	0.07	0.03	2.08	0.04	0.19	0.15	1.24	0.22
Video Narratives Only	0.08	0.04	1.95	0.05	0.10	0.03	2.94	0.00	0.24	0.11	2.24	0.03

By participant political knowledge (canvass)

In the first post-treatment survey, we asked five political knowledge questions on the length of a presidential term, the length of a Senate term, the purpose of Medicare, the Chief Justice of the Supreme Court, and on which program the US spends the least. We then coded individuals into how many questions they answered correctly out of 5.

For sample size considerations, we group together all respondents who answered 0, 1, or 2 questions correctly.

Table OA55: Heterogeneous treatment effects by political knowledge

	5/5				4/5				3/5				<2/5			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Participants' and Video Narratives	0.08	0.05	1.61	0.11	0.11	0.04	2.82	0	0.11	0.07	1.44	0.15	0.12	0.06	1.92	0.06
Video Narratives Only	0.07	0.05	1.60	0.11	0.13	0.04	3.20	0	0.09	0.05	1.72	0.09	0.13	0.07	2.02	0.04

By participant party identification (canvass)

In the canvass, 407 conversations and one week surveys were completed by voters who self-identified in the baseline survey as Democrats; 295 by self-identified Republicans; and 342 by self-identified Independents (including party leaners).

Table OA56: Heterogeneous treatment effects by voter party identification

	Democrats				Republicans				Independents			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Participants' and Video Narratives	0.12	0.04	2.94	0	0.15	0.06	2.57	0.01	0.04	0.04	0.90	0.37
Video Narratives Only	0.12	0.03	3.45	0	0.13	0.05	2.43	0.02	0.06	0.04	1.33	0.18

By participant political knowledge (phone)

For sample size considerations, we group together all respondents who answered 0, 1, or 2 questions correctly.

Table OA57: Heterogeneous treatment effects by political knowledge

	5/5				4/5				3/5				<2/5			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Phone	0.05	0.03	2	0.05	0.02	0.02	0.64	0.52	0.07	0.03	2.32	0.02	0.07	0.04	1.64	0.1

By participant party identification (phone)

In the phone, 727 conversations and one week surveys were completed by voters who self-identified in the baseline survey as Democrats; 605 by self-identified Republicans; and 611 by self-identified Independents (including party leaners).

Table OA58: Heterogeneous treatment effects by voter party identification

	Democrats				Republicans				Independents			
	Effect	SE	t.stat	p	Effect	SE	t.stat	p	Effect	SE	t.stat	p
Phone	0.04	0.02	2.1	0.04	0.03	0.03	0.91	0.36	0.06	0.03	2.25	0.02

Estimates for Dichotomized Policy Items

It may be difficult for readers to interpret the magnitude of an effect presented in terms of standard deviation change. We therefore take two, non-pre-registered approaches to help communicate the substantive size of our estimates.

Strong Support - Canvass (Experiment 2)

First, one way to make the results more interpretable is to examine treatment effects on whether participants said they strongly supported the policies asked about in the surveys. This attempts to recreate how participants might vote on each proposal if faced with a ballot measure or was deciding between candidates who differ on their transgender non-discrimination proposals. Note that we did not pre-specify this benchmarking procedure. We use this to illustrate the magnitude of our findings. In particular, we report results on the individual dichotimized items from the first post-treatment survey, set to 1 if an individual took a strong position on the supportive side of the issue and 0 otherwise. For these analyes for Experiment 2, we combine the two treatments for simplicity and statistical power.

The results on the individual dichotomized items are as follows:

Table OA59: At max at t1 (ITTs)

Variable Name	Effect	SE	t.stat	p
Non-Discrimination Law (Before Video)	0.025	0.022	1.143	0.253
Non-Discrimination Law (After Video)	0.081	0.022	3.588	0.000
Protect From Firing For Being Trans	0.062	0.022	2.758	0.006
Must Use Bathroom Or Locker Matching Sex At Birth (Reverse Coded)	0.033	0.025	1.309	0.191
Concerned About Sex Predators and Non-Discrim Law (Reverse Coded)	0.023	0.021	1.081	0.280
Trans People Should Not Be Public School Teachers (Reverse Coded)	0.014	0.025	0.572	0.567

Any Support - Canvass (Experiment 2)

Second, we also conduct a version of this benchmarking where we dichotomize each variable to record whether participants registered any support (not only strong agreement). These new dichotomized variables are coded to 1 if a participant agreed at all with the policy and 0 otherwise (indicating stated indifference or opposition). The results on the individual items dichotimized in this manner are as follows:

Table OA60: Agreement at all at t1 (ITTs)

Variable Name	Effect	SE	t.stat	p
Non-Discrimination Law (Before Video)	-0.010	0.018	-0.550	0.583
Non-Discrimination Law (After Video)	0.027	0.021	1.281	0.201
Protect From Firing For Being Trans	0.025	0.019	1.316	0.188
Must Use Bathroom Or Locker Matching Sex At Birth (Reverse Coded)	0.024	0.025	0.942	0.347
Concerned About Sex Predators and Non-Discrim Law (Reverse Coded)	0.063	0.020	3.180	0.002
Trans People Should Not Be Public School Teachers (Reverse Coded)	0.042	0.022	1.920	0.055

Strong Support - Phone (Experiment 3)

The results on the individual dichotomized items when dichotomized to capture strong support only for the phone experiment (Experiment 3) are as follows:

Table OA61: At max at t1 (ITTs)

Variable Name	Effect	SE	t.stat	p
Non-Discrimination Law (Before Video)	0.023	0.016	1.468	0.142
Non-Discrimination Law (After Video)	0.020	0.016	1.256	0.209
Protect From Firing For Being Trans	0.043	0.016	2.650	0.008
Must Use Bathroom Or Locker Matching Sex At Birth (Reverse Coded)	-0.003	0.017	-0.169	0.866
Concerned About Sex Predators and Non-Discrim Law (Reverse Coded)	-0.011	0.014	-0.740	0.460
Trans People Should Not Be Public School Teachers (Reverse Coded)	-0.010	0.018	-0.574	0.566

Any Support - Phone (Experiment 3)

The results on the individual dichotomized items when dichotomized to capture any support for the phone experiment (Experiment 3) are as follows:

Table OA62: Agreement at all at t1 (ITTs)

Variable Name	Effect	SE	t.stat	p
Non-Discrimination Law (Before Video)	-0.032	0.014	-2.331	0.020
Non-Discrimination Law (After Video)	-0.014	0.015	-0.929	0.353
Protect From Firing For Being Trans	0.000	0.015	-0.013	0.990
Must Use Bathroom Or Locker Matching Sex At Birth (Reverse Coded)	0.019	0.018	1.065	0.287
Concerned About Sex Predators and Non-Discrim Law (Reverse Coded)	0.012	0.014	0.853	0.394
Trans People Should Not Be Public School Teachers (Reverse Coded)	0.015	0.016	0.931	0.352

References

[1] David E Broockman, Joshua L Kalla, and Jasjeet S Sekhon. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis*, 25(4):435–464, 2017.

[2] Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation*. WW Norton, 2012.

[3] Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46, 2012.

[4] Richard E. Petty, Curtis P. Haugtvedt, and Stephen M. Smith. Elaboration as a determinant of attitude strength: Creating attitudes that are persistent, resistant, and predictive of behavior. In R.E. Petty and J.A. Krosnick, editors, *Attitude strength: Antecedents and consequences*, pages 93–130. Lawrence Erlbaum Associates, Inc., 1995.