Supplementary Materials for Durably reducing transphobia: a field experiment on door-to-door canvassing

David E. Broockman,[1]* Joshua L. Kalla[2]

[1] Graduate School of Business, Stanford University

[2] Department of Political Science, University of California, Berkeley

*Correspondence to: dbroockman@stanford.edu.

This PDF file includes:

- Materials and Methods
- Fig. S1
- Tables S1 - S25

Materials and Methods

The full replication code that produces this report will be available at http://dx.doi.org/10.7910/DVN/WKR39N.

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 Miami-Dade County. To recruit voters to these surveys, the Los Angeles LGBT Center 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 in Miami-Dade County, Florida, USA. We invited these voters to the baseline survey by mail. The survey was called the "2015 Miami-Dade 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 35550 households that contained individual logins for 68378 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 incentives of $5, $8, $10, $15, $20, or $30 for completing the baseline survey and $5, $10, or $20 for completing each follow-up survey. We randomly assigned households to these incentive amounts in advance. Voters could opt to receive these incentives in the form of cash sent in the mail within 2 weeks, an Amazon.com gift card, or a donation to one of four charities (the Red Cross, the United Way, Big Brothers and Big Sisters of Greater Miami, or the American Cancer Society). Most respondents chose to receive cash in the mail.

Random Assignment of Households

1825 voters completed the baseline survey and provided a valid email address. We randomly assigned half of baseline survey respondents to be targeted with the treatment and half to be targeted with the placebo.

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 (*33*). All analyses adjust standard errors to account for this clustered assignment (see details below) (*33*).

The household-level clustered random assignment took place within blocks of two households. These blocks were formed by matching households with similar household-level-average values of baseline covariates. Within each block, one household (cluster) was assigned to treatment and one to placebo. This pre-treatment blocking reduces the chance of imbalance between conditions and improves precision *(33)*.

The following STATA code was used for random assignment.

```
set seed 54811
set sortseed 94786

//Average the scale at the HH level for blocking at the HH level
preserve
```

```
bysort hh_id: egen scale_t0_hh_avg = mean(scale_for_blocking_t0)
keep hh_id scale_t0_hh_avg
duplicates drop hh_id, force

//Generate pairs of two HHs
sort scale_t0_hh_avg
egen block_ind=seq(), block(2)
gen rand=runiform()
sort block rand
by block: gen treat_ind=_n==1

keep treat_ind block_ind hh_id
tempfile ind_rand
save `ind_rand'
restore

merge m:1 hh_id using `ind_rand', nogen
```

Random Assignment of Turfs

On the day of each canvass, groups of households were formed into "turfs" by the staff at the Los Angeles LGBT Center and SAVE. "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. Data-quality checks conducted after the canvass ensured that canvassers all properly canvassed the assigned doors within their turf.

Placebo Design for Delivering Intervention

Canvassers attempted to have a conversation about recycling with voters in the placebo group and a conversation containing the intervention with voters in the treatment group. This placebo-controlled experimental design (*19*) is common in studies of door-to-door canvassing interventions (e.g., (*34, 35*)) and field experiments more generally (e.g., (*36, 37*)). Nickerson (*19*) 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. (*38*) 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 about recycling 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 (*19, 33, 34*).

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 implemeting 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 transgender issues or recycling 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 (*19*). Fig. S1 below provides an overview of this implementation.

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 either condition. These follow-up surveys began 3 days, 3 weeks, 6 weeks, and 3 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. On the day the survey opened, and if the voter provided a cell phone number, we also sent a text message reminder concurrently with the email invitation. For the 3-day survey only, if voters did not complete the survey after 7 days, we called voters personally to ask them to complete the survey and sent them a personally signed letter asking them to do so. All follow-up surveys were open for approximately 2 weeks.

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 (*33*).

Additional Survey Details

The survey was called the 2015 Miami-Dade Opinion Study, conducted by the University of California, Berkeley. 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 conservative neighborhoods in Miami-Dade county, as chosen by staff at the Los Angeles LGBT Center and SAVE. These neighborhoods were picked on the basis of their historic opposition to LGBT-inclusive ballot measures and walkability. Voters were recruited from this population by mail we sent to their household using the mail vendor Snappack Mail (http://www.snappackmail.com). Multiple voters sharing the same household were invited to participate in the survey, up to 3 per household.

The sampling frame was this population with a few exclusions. If more than 3 voters shared a household, the 3 invited to participate were randomly sampled. Per standard practice among political canvassing efforts, several types of households were excluded from the initial sample for fear that they represented low data quality or homes that could not be canvassed: households with a mailing address outside of Florida, households where the mailing address did not match the address where the voter was registered to vote, households with more than 5 registered voters, and households in apartment buildings. All other voters were mailed an invitation to participate in the survey.

Table S20 shows how the representativeness of those who responded to the survey differ from those mailed an invitation to participate in the survey.

Table S25 gives the AAPOR standard definition response rates to each survey.

Two different Miami-Dade neighborhoods were mailed on or around May 6, 2015 and June 4, 2015. Canvassing for the first neighborhood occurred on June 6, 10, and 17, 2015. Canvassing for the second neighborhood occurred on June 20, 24, and 27, 2015. Those canvassed were then invited to participate in the follow-up surveys exactly 3 days, 3 weeks, 6 weeks, and 3 months after they were contacted at the door.

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.

Pre-Analysis Plans

We filed three pre-analysis plans with EGAP (id 20150707AA at www.egap.org). These are available in the replication data as well.

The first pre-analysis plan notes that it was filed after data from the three-day survey was collected but before it was analyzed with the treatment indicator. We sought advice from multiple colleagues about estimation procedures, and this feedback did not come back in time for us to file a plan in advance. The main difference between the plan as filed and the paper is that we examine the outcomes related to transgender prejudice and the law separately. For completeness we also report the results for the original index of all outcomes as we had pre-registered it in the Supplementary Materials, the "all.dvs" index (see tables S1-S4), which we referred to as "main.dv" in the plan; the two indicies we report in the paper (trans.prejudice and miami_trans_law) split the outcomes in the "all.dvs"/"main.dv" index into the law and transgender items. Splitting the items we pre-registered as the main DVs in this manner allowed us to be transparent about the fact that, as Fig. 2 in the paper shows, there was a null effect on the outcomes related to the law initially.

The second pre-analysis plan was filed to note that we had split the outcomes in this way and planned to track persistence on each separately.

The third pre-analysis plan was filed to register our expectations about adding the definition to the law item potentially revealing an effect, as providing this definition better approximated the circumstances under which people vote on the law.

Outcomes

The survey included dozens of political, social, and cultural questions, only some of which were related to transgender prejudice. Prior to analyzing the data with the treatment indicator we submitted a pre-analysis plan to the Evidence in Governance and Politics (EGAP) trial registry to indicate which items constituted experimental outcomes. Below we list these items and give their full text.

Rationale for Outcome Selection

For our outcomes, we mainly drew items from existing batteries and made minor modifications *(39-42)*. The literature on measuring transgender stigma is still in its infancy, and many existing batteries were designed for

specific situations or used outdated terminology. Unlike the literature on anti-black prejudice, there are also no widely accepted batteries for measuring anti-transgender prejudice yet. Although we would have preferred to rely on such a scale, our goal here was not to create such a scale. Instead, conceptualizing prejudice as we define it early in the paper – broadly as negative attitudes towards an outgroup – we therefore included items from several different existing batteries to create an index that would measure many different pervasive negative attitudes about transgender people other scholars have identified.

We take no position on whether these items all reflect a single latent dimension; the literature does, however, indicate they all represent negative attitudes towards transgender people. We pre-registered using factor analysis to combine the items instead of a simple additive index because we anticipated that some items would contain more measurement error than others but did not know which ex ante; factor analysis provided a natural way to weight items that contained less noise.

We added two additional items in the 3-week survey based on reviewing advertisements and arguments that had been used in the past to support excluding transgender people from non-discrimination laws. We thought it was important to understand whether the treatment reduced the stereotypes that opponents of non-discrimination protections capitalize on so added these.

We avoided using the term "transgender" in nearly all of the items without defining it for fear that many subjects in the placebo group would not be familiar with the term (potentially being more familiar with other, derogatory terms for this group). Two exceptions where we did use the term "transgender" without defining it are the item about the law and the feeling thermometer.

For the item about the law, we did not define the term transgender because ballot language does not define it either. However, as political campaigns often do define the term in advance of public votes, we added the definition starting in the six-week survey to see what the effects might look like if we attempted to approximate these circumstances (there was no active political campaign in Miami at the time).

For the feeling thermometer, we did not define the term because all the other feeling thermometer items were very brief (e.g., "President Obama"); we thought providing a definition for this one feeling thermometer would have called attention to this item and potentially raised suspicion about the survey's connection to the treatment.

Tables S1-S4 show the effects on all items separately and on the indices we create of all items.

Items Capturing Primary Outcomes Appearing on All Surveys

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 3-day survey is survey 1;
- the 3-week survey is survey 2;
- the 6-week survey is survey 3, and;
- the 3-month survey is survey 4.

Item names appear in bold next to each item. For the remainder of the paper we will refer to these items by these bolded names.

- **miami_trans_law_t#**: Miami-Dade county recently passed a law that prohibits discrimination in housing, employment and public accommodations based on gender identity and expression, a category that includes transgender men and women [for **miami_trans_law_withdef_t#**, adds the phrase here, "(people who were designated one gender at birth, but now identify as a different gender)."]. Do you favor or oppose this new law?

- **miami_trans_law2_t#**: Some people say it's important to protect transgender people from discrimination in housing and employment. Other people have concerns about society becoming too accepting of transgender people, and do not want transgender people included in our non-discrimination law. What do you think? Do you agree or disagree that Miami law should protect transgender people [for **miami_trans_law2_withdef_t#**, adds the phrase here, "(people who were designated one gender at birth, but now identify as a different gender)."] from discrimination?
- **trans_law_post_ad_t3**: On the third survey only, we asked about the law again after showing an opposition video: "We've mentioned that Miami-Dade county recently passed a law that prohibits discrimination in housing, employment and public accommodations based on gender identity and expression, a category that includes transgender men and women (people who were designated one gender at birth, but now identify as a different gender). After seeing the video, do you favor or oppose this new law?" This item does not appear in any of the scales.
- **therm_trans_t#**: Feeling thermometer towards trans people (0-100) (From *(39)*)
- **gender_norm_abnormal_t#**: A man who identifies as a woman is psychologically abnormal (Modification from (*40, 41*)).
- **gender_norm_moral_t#**: It is morally wrong for a man to present himself as a woman in public (Modification from (*40*)).
- **gender_norm_trans_moral_wrong_t#**: Saying you are a gender that is different than the one you were born with is morally wrong (Modification from (*40, 41*)).
- **gender_norm_sexchange_t#**: I would support a friend choosing to have a sex change (Modification from (*42*)).

Starting with the second post-treatment survey, we also asked the below two items. These items were composed and added after examining the results from the first post-treatment survey. An amended pre-analysis plan was submitted to EGAP to reflect the addition of these two questions.

- **trans_teacher_t#**: Transgender women (people who identify as women but were designated male at birth) should be allowed to serve as public school teachers.
- **trans_bathroom_t#**: It would be wrong to allow a transgender woman (a person who identifies as a woman but was designated male at birth) to use the woman's restroom (Modification from (*40*)).

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 (*33*).

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

- **all.dvs.t#**: An index of all primary outcomes, created to test the omnibus hypothesis that the treatment had any effects.
- **trans.tolerance.dv.t#**: An index of outcomes from all.dvs.t# capturing acceptance and tolerance towards – instead of prejudice and stigma towards – transgender people; this is all primary outcomes except for the two items about the law.
- **miami_trans_law_t#_avg**: An index of the two items about Miami's law protecting transgender people from discrimination in housing, employment, and public accomodations. Because these items were Likert scales with the same number of points and there were only two of them, the index is a simple average of the two items.
- **gender_nonconformity_t#**: Distinct from measuring transgender tolerance, we also constructed a scale capturing gender norms. As stated in our pre-analysis plan, effects on this measure represented a secondary hypothesis and this index is not a primary outcome of interest. (However, there do appear to be effects on this measure as well, but we do not discuss these in the main text due to space constraints.)

Pre-Specified Index of All Primary Outcomes In the pre-analysis plan we indicated that the main hypothesis test will be whether an index of all primary outcomes shows statistical significance. We list these outcomes for each survey below. (In the pre-analysis plan we called this "main.dv.")

Note that we code the scale such that larger, more positive values indicate more tolerance and less prejudice.

```
all.dv.names.t1 <- c('miami_trans_law_t1', 'miami_trans_law2_t1', 'therm_trans_t1',
                     'gender_norm_sexchange_t1', 'gender_norm_moral_t1',
                     'gender_norm_abnormal_t1', 'gender_norm_trans_moral_wrong_t1')

all.dv.names.t2 <- c('miami_trans_law_t2', 'miami_trans_law2_t2', 'therm_trans_t2',
                     'gender_norm_sexchange_t2', 'gender_norm_moral_t2',
                     'gender_norm_abnormal_t2', 'gender_norm_trans_moral_wrong_t2')

all.dv.names.t3 <- c('miami_trans_law_withdef_t3', 'miami_trans_law2_withdef_t3',
                     'therm_trans_t3', 'gender_norm_sexchange_t3', 'gender_norm_moral_t3',
                     'gender_norm_abnormal_t3','gender_norm_trans_moral_wrong_t3')

all.dv.names.t4 <- c('miami_trans_law_withdef_t4', 'miami_trans_law2_withdef_t4',
                     'therm_trans_t4', 'gender_norm_sexchange_t4', 'gender_norm_moral_t4',
                     'gender_norm_abnormal_t4', 'gender_norm_trans_moral_wrong_t4')
```

Transgender Tolerance Index After analyzing the 3 day survey results, we found a clear difference between the treatment effects on the two questions about the law and the treatment effects on the items which captured tolerance and prejudice. We anticipated wanting to discuss these themes separately in the paper. Therefore, we decided to split the items in the omnibus index into two indicies: first, we kept track of the two law items in one index (miami_trans_law_t#_avg); second, we created a second index with all the remaining main outcome items, which focused on the questions related to prejudice (trans_prejudice_dv_t#). These changes were specified in an amended pre-analysis plan prior to finishing data collection for or analyzing any data from the 3-week survey.

Beginning with the 3-week survey we also added two items to the transgender tolerance index that appeared on the 3-week survey for the first time, one about whether transgender women should be permitted to serve as public school teachers and the other about whether transgender women should be allowed to use the woman's restroom (see **trans_teacher_t#** and **trans_bathroom_t#** defined above). These items were composed and added prior to the analysis of the 3-week or the 3-day survey. See "Rationale for Outcome Selection."

Note that we code the scale such that larger, more positive values indicate more tolerance and less prejudice.

```
trans.tolerance.dvs.t0 <- c('therm_trans_t0', 'gender_norms_sexchange_t0',
                            'gender_norms_moral_t0', 'gender_norms_abnormal_t0')

trans.tolerance.dvs.t1 <- c('therm_trans_t1', 'gender_norm_sexchange_t1',
                            'gender_norm_moral_t1', 'gender_norm_abnormal_t1',
                            'gender_norm_trans_moral_wrong_t1')

trans.tolerance.dvs.t2 <- c('therm_trans_t2', 'gender_norm_sexchange_t2',
                            'gender_norm_moral_t2', 'gender_norm_abnormal_t2',
                            'gender_norm_trans_moral_wrong_t2',
                            'trans_teacher_t2', 'trans_bathroom_t2')

trans.tolerance.dvs.t3 <- c('therm_trans_t3', 'gender_norm_sexchange_t3',
                            'gender_norm_moral_t3', 'gender_norm_abnormal_t3',
```

```
                              'gender_norm_trans_moral_wrong_t3',
                              'trans_teacher_t3', 'trans_bathroom_t3')

trans.tolerance.dvs.t4 <- c('therm_trans_t4', 'gender_norm_sexchange_t4',
                              'gender_norm_moral_t4', 'gender_norm_abnormal_t4',
                              'gender_norm_trans_moral_wrong_t4',
                              'trans_teacher_t4', 'trans_bathroom_t4')
```

Law Items Index As mentioned above, we created two separate indicies that differentiated between the prejudice questions and the questions about the law. Below are the indices created for each survey wave about the law.

```
trans.law.dvs.t0 <- c('miami_trans_law_t0', 'miami_trans_law2_t0')
trans.law.dvs.t1 <- c('miami_trans_law_t1', 'miami_trans_law2_t1')
trans.law.dvs.t2 <- c('miami_trans_law_t2', 'miami_trans_law2_t2')
# Note: Beginning with t3, the definition was added.
trans.law.dvs.t3 <- c('miami_trans_law_withdef_t3', 'miami_trans_law2_withdef_t3')
trans.law.dvs.t4 <- c('miami_trans_law_withdef_t4', 'miami_trans_law2_withdef_t4')
```

Secondary Outcome: Gender Non-Conformity Index We were also interested in whether the treatment would change attitudes towards gender non-conformity more generally (individuals can be gender non-conforming without being transgender). We pre-registered this hypothesis as a secondary hypothesis and not the main hypothesis of interest. However, we present the results below for completeness. Our measure of gender non-conformity consisted of an index built from the following items:

- **gender_norm_looks_t#**: To keep children from being confused, it's better when men look and act like men, and women look and act like women.
- **gender_norm_rights_t#**: Men and women should have equal rights, but men and women are not the same; it's normal for men to act like men, and women to act like women.
- **gender_norm_dress_t#**: Men should dress like men and women should dress like women. (This item was only added beginning with t2, the 3 week follow-up survey.)

```
gender.nonconformity.t0 <- c('gender_norm_looks_t0', 'gender_norm_rights_t0')
gender.nonconformity.t1 <- c('gender_norm_looks_t1', 'gender_norm_rights_t1')
# Note: Beginning with t2, an additional item was added to the measure.
gender.nonconformity.t2 <- c('gender_norm_looks_t2', 'gender_norm_rights_t2',
                              'gender_norm_dress_t2')
gender.nonconformity.t3 <- c('gender_norm_looks_t3', 'gender_norm_rights_t3',
                              'gender_norm_dress_t3')
gender.nonconformity.t4 <- c('gender_norm_looks_t4', 'gender_norm_rights_t4',
                              'gender_norm_dress_t4')
```

Reverse Coded Items

We reverse the reverse-coded items, such that positive values on all variables would indicate more tolerant attitudes and success for the intervention.

Procedue for Combining Outcomes into Indices

We pre-specified the procedure below to calculate each index. In the interest of transparency we provide the full code.

Note that we code all indicies such that higher values on the indices indicate more tolerance and success of the intervention.

```
# Compute factor analysis outcome
compute.factor.dv <- function(dv.names, respondent.booleans, print.loadings = TRUE){
  responders <- data[respondent.booleans,]

  # Factor analysis
  factor.obj <- princomp(responders[, dv.names], cor=TRUE)
  if(print.loadings) print(loadings(factor.obj))
  dv <- as.vector(factor.obj$scores[,1])

  # More positive values on the factor should indicate more tolerance; reverse otherwise.
  if(cor(dv, responders$miami_trans_law_t0, use="complete.obs") < 0) dv <- -1 * dv

  # Put in the order of the main data frame
  dv.in.order <- dv[match(data$id, responders$id)]

  # Rescale to mean 0 sd 1 in placebo group; treatment effects can then be interpreted
  # as the effect in standard deviations the treatment would have among an untreated
  # population.
  dv.in.order <- (dv.in.order - mean(dv.in.order[!data$treat_ind], na.rm=TRUE)) /
    sd(dv.in.order[!data$treat_ind], na.rm=TRUE)

  return(as.vector(dv.in.order))
}
```

Estimation Procedures

Contact Rate

501 voters identified themselves at the door after the initial greeting that did not differ by condition. Canvassers then either began an intervention conversation or a placebo conversation. Of the 246 voters who identified themselves at their doors in the treatment group, 192 began the conversation and at least described their initial view on the law to the canvasser, rather than refusing to talk at all after identifying themselves. On the other hand, the treatment was inadvertendly delivered to 11 individuals in the placebo group due to canvasser error. Consistent with our pre-analysis plan, we report estimated complier average causal effects for the intervention under the assumptions 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 (*33*). Reporting these point estimates does not change the experimental comparison we conduct, but does increase point estimates slightly to account for the measurement error in the treatment indicator. Relaxing these assumptions and reporting average intent-to-treat effects would decrease the point estimates slightly but would not change any t-ratios or p-values (*33*).

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.

Complier Average Causal Effect Estimation

To estimate complier average causal effects, we rely on ordinary least squares (OLS) with cluster-robust standard errors, clustering on household and residualizing using pre-treatment covariates from the baseline survey and voter list. This procedure and these covariates were pre-specified in advance and produce unbiased estimates of causal effects (*33*). We also adjust for the contact rate, as described above. Finally, we implement Olken's rejection rule (*43*) as specified in our pre-analysis plan, which denoted the bottom 1% and top 4% of the sampling distribution as the rejection region (instead of either the bottom 2.5% and top 2.5% or only top 5% as is conventional). We report p-values for this region and otherwise report "n.s." (*43*). (The only result this rule affects is the 3-month effect on the law index.)

```
t0.covariate.names <- c('miami_trans_law_t0', 'miami_trans_law2_t0', 'therm_trans_t0',
'gender_norms_sexchange_t0', 'gender_norms_moral_t0', 'gender_norms_abnormal_t0',
'ssm_t0', 'therm_obama_t0', 'therm_gay_t0','vf_democrat', 'ideology_t0',
'religious_t0', 'exposure_gay_t0', 'exposure_trans_t0', 'pid_t0', 'sdo_scale',
'gender_norm_daugher_t0', 'gender_norm_looks_t0',
'gender_norm_rights_t0', 'therm_afams_t0', 'vf_female', 'vf_hispanic',
'vf_black', 'vf_age', 'survey_language_es', 'cluster_level_t0_scale_mean')
x <- data[,c(t0.covariate.names)]
x <- as.matrix(x, dimnames = list(NULL, names(x)))

# Function to compute clustered standard errors, from Mahmood Arai.
cl <- function(fm, cluster){
  M <- length(unique(cluster))
  N <- length(cluster)
  K <- fm$rank
  dfc <- (M/(M-1))*((N-1)/(N-K))
  uj  <- apply(estfun(fm), 2, function(x) tapply(x, cluster, sum))
  vcovCL <- dfc*sandwich(fm, meat=crossprod(uj)/N)
  coeftest(fm, vcovCL)
}

# Function to extract the average treatment effect from OLS with clustered SEs.
est.ate <- function(dv, include.obs = NULL, include.covariates = TRUE){
  if(is.null(include.obs)){
    include.obs <- !is.na(dv)
  }
  include.obs <- which(include.obs & !is.na(dv))

  if(include.covariates) lm.obj <- lm(dv[include.obs] ~ data$treat_ind[include.obs] +
                                        x[include.obs,])
  if(!include.covariates) lm.obj <- lm(dv[include.obs] ~ data$treat_ind[include.obs])

  # Calculate cluster-robust standard errors.
  result <- cl(lm.obj, data$hh_id[include.obs])[2,]

  # Adjust point estimate and standard error for contact rate in subsample.
  itt_d <- lm(treatment.delivered ~ treat_ind, data[include.obs,])$coefficients[2]
  result[1:2] <- result[1:2] / itt_d

  # Per pre-analysis plan, rejection region is top 4% and bottom 1% of sampling distribution,
  # so p-values are reported for this region; otherwise, we write "n.s."
  # See Olken (*43*) , page 70, footnote 5.
  # Note that cl() returns two-tailed p-values that must be converted to one-tailed.
```

```
  # Note that all DVs were recoded such that higher values indicated more tolerance.
  result[4] <- result[4] / 2 # p-value corresponds to mass under one side of distribution.
  rejection.region <- ((result[4] < .04 & result[1] > 0) | # Significant positive result.
                         (result[4] < .01 & result[1] < 0)) # Significant negative result.
  result <- round(result, 3)
  if(rejection.region) result[4] <- paste0(as.character(result[4]), "*")
  if(!rejection.region) result[4] <- "*n.s.*"
  if(result[4] == "0*") result[4] <- "0.000*" # Indicate precision of 0 p-value.
  names(result)[4] <- "*p*"
  return(result)
}
```

Heterogenous Treatment Effects

We searched for heterogenous treatment effects by computing a residualized trans.tolerance.dv.t1 among t1 responders (residualizing using the pre-specified covariates but not the treatment indicator) then testing whether the treatment effect on this residualized index was larger for some subgroups. (Residualizing first makes sure the coefficients on the covariates do not differ within subgroups when we compute estimates within subgroups.) In our pre-analysis plan, we specified three predictions about treatment effect heterogeneity by subject attributes:

1. Democrats will be more treatment-responsive than Republicans as a result of partisan cues being more aligned,
2. Subjects higher on the baseline support scale will be more treatment-responsive as a result of being more open to outgroups in general, and
3. Subjects higher in need for cognition will show larger and longer-lasting effects (this scale was only included on the six week survey, so we view this analysis as more exploratory given that it was not measured pre-treatment).

We found evidence for none of these patterns. In the last subsection we also show evidence of insignificant heterogenous treatment effects for a broader set of covariates.

Note that we present tests for the presence of heterogeneous treatment effects for completeness, but the fact that we fail to find such effects may reflect low power meaning we are unable to reject a false null; the below tests should not be construed as our accepting the null of no difference.

By Party Registration We analyze the data two ways. In Table S6, we present the conditional average treatment effect among Democrats, Independents, and Republicans on the trans.tolerance.dv.t1 factor. In Table S7, we present the interaction effect between binary variables for Democrats and Independents and the treatment indicator on the same factor (making Republicans the base category), after having been residualized on the above specified covariates. We find no evidence of heterogeneous treatment effects. (The only hint of a pattern is that Republicans are more affected than Independents.)

Again, recall that because the independent variable has already been residualized, the base terms do not reflect the baseline difference between the groups (e.g., Republicans are indeed less supportive at baseline).

By Baseline Support We analyze the data two ways. In Table S8, we present the conditional average treatment effect among individuals below the mean on the baseline support factor and above the mean on the trans.tolerance.dv.t1 factor. In Table S9, we present the interaction effect between a continuous variable for baseline support and the treatment indicator on the same factor, after having been residualized on the above specified covariates. We find no evidence of heterogeneous treatment effects.

Note that the correlation between the baseline and the outcome is weak because we compute heterogenous treatment effects on a residualized version of the outcome.

By Need for Cognition We analyze the data two ways. In Table S10, we present the conditional average treatment effect among individuals below the mean on the Need for Cognition factor and above the mean on the trans.tolerance.dv.t1 factor. In Table S11, we present the interaction effect between a continuous variable for Need for Cognition and the treatment indicator on the same factor, after having been residualized on the above specified covariates. We find no evidence of heterogeneous treatment effects.

Recall that the Need for Cognition item was only asked on the third post-treatment survey, not the baseline survey. This means additional caution is warranted because the item was asked post-treatment; it is also not available for individuals who did not complete the t3 survey.

Machine Learning Algorithm for Other Subgroups There are many other covariates available from the voter file and on our survey. In order to avoid a post hoc hypothesis testing exercise that would likely lead us to find at least one such covariate to predict the treatment effect purely by chance, we also implemented a machine learning procedure designed to automatically search for the existence of any robust heterogenous treatment effects (*44*). The procedure relies on the Lasso, a variable selection algorithm. The results return only the intercept, indicating that none of the covariates predict the treatment effect robustly; the treatment is broadly effective among subgroups. We take this result as propitious for the generalizability of the findings.

```
# Residualize the dependent variable, then transform it per (*44*).
t1.resid <- summary(lm(data$trans.tolerance.dv.t1 ~ x))$residuals
data$t1.resid[as.numeric(names(t1.resid))] <- t1.resid # Maps residuals back into data.
data$transformed.outcome <- with(data, t1.resid * (treat_ind - .5) / .25)

# Vectors describing rows where the outcome was observed.
include.obs <- !is.na(data$transformed.outcome)
x.subset <- cbind(data$treat_ind[include.obs], x[include.obs,])
transformed.outcome <- data$transformed.outcome[include.obs]

# Lasso.
cvfit.lasso <- cv.glmnet(x.subset, transformed.outcome, alpha = 1)
coef(cvfit.lasso, s = "lambda.min")
```

```
## 28 x 1 sparse Matrix of class "dgCMatrix"
##                                       1
## (Intercept)                 0.2079517
##                             .
## miami_trans_law_t0          .
## miami_trans_law2_t0         .
## therm_trans_t0              .
## gender_norms_sexchange_t0   .
## gender_norms_moral_t0       .
## gender_norms_abnormal_t0    .
## ssm_t0                      .
## therm_obama_t0              .
## therm_gay_t0                .
## vf_democrat                 .
## ideology_t0                 .
## religious_t0                .
## exposure_gay_t0             .
## exposure_trans_t0           .
## pid_t0                      .
## sdo_scale                   .
## gender_norm_daugher_t0      .
```

```
## gender_norm_looks_t0        .
## gender_norm_rights_t0       .
## therm_afams_t0              .
## vf_female                   .
## vf_hispanic                 .
## vf_black                    .
## vf_age                      .
## survey_language_es          .
## cluster_level_t0_scale_mean .
```

Tests of Design Assumptions

Covariate Balance among All Subjects, Compliers, and Reporters

Tables S12-S17 demonstrate that balance on pre-treatment observable attributes is maintained among the original universe of pre-survey respondents randomized to each group, the subsample that was canvassed, and the subsample that was both canvassed and successfully re-interviewed. Each table shows the mean value for the covariate under treatment and placebo as well as the p-value from a t-test of the difference in means. Table S12 considers all voters who were randomly assigned after having taken the pre-survey (all subjects); Table S13 considers all voters who were successfully contacted (compliers); Tables S14-S17 considers all voters who responded to the first through fourth post-survey (reporters).

Survey Attrition

An important design assumption is that the treatment does not affect the composition of the individuals who take each follow-up survey (*33*).

Test of Average Differential Attrition Table S18 shows the counts of respondents to each survey wave by experimental condition.

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) *(33, 45)*. 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. For this test, we compare the F-statistic in the observed data to a reference distribution of F-statistics that would be observed under the sharp null hypothesis. This procedure was pre-specified in our pre-analysis plan and is standard practice *(33)*.

Based on the results presented in Table S19, there does not appear to be evidence of asymmetrical attrition.

Representativeness

Table S20 demonstrates that the subjects remained representative of the starting universe at each stage.

Placebo Tests On Additional Survey Items

A reviewer helpfully noted that we can use the survey items unrelated to prejudice, discrimination, or LGBT people as placebo tests where none of these items change in the manner that the transgender attitude items change. This test was not included in our pre-analysis plan, but is reported in Table S21 for some of the distractor items. There were hundreds of items across all the follow-up surveys. In Table S21 we show that thermometer ratings of President Obama, Jeb Bush, marijuana users, police officers, firefighters, and former Prime Minister Mir-Hossein Mousavi were similar across conditions. Note that not all of these items were asked on all of the follow-up surveys. We report all results for which the question was asked and all the items we examined.

Further Explanation of and Replication Code for Results in Main Text

Results In Main Text

For specificity and transparency, below we list the replication code that produces every numerical or statistical result described in the main text. We also elaborate on some of the results that space constraints prevented us from explaining further in the main text.

- "Conversations lasted around 10 minutes on average.""

```
mean(data$canvass_minutes[data$treatment.delivered], na.rm=TRUE)
```

```
## [1] 10.54187
```

- "In addition, geographic clusters of respondents were randomly assigned to receive the intervention from either a transgender (n=15) or non-transgender (n=41) volunteer.""

```
nrow(aggregate(data$canvasser_trans[data$canvasser_trans==1],
list(data$canvasser_id[data$canvasser_trans==1]), unique))
```

```
## [1] 15
```

```
nrow(aggregate(data$canvasser_trans[data$canvasser_trans==0],
list(data$canvasser_id[data$canvasser_trans==0]), unique))
```

```
## [1] 41
```

- "To measure outcomes, we conducted follow-up surveys of individuals who had come to their doors in either condition (n=501)."

```
table(data$contacted)
```

```
##
##   1
## 501
```

- "These surveys began three days (n = 429), three weeks (n = 399), six weeks (n = 401), and three months (n = 385) after the canvassing took place."

```
table(data$respondent_t1)
```

```
##
##   0   1
##  72 429
```

```
table(data$respondent_t2)
```

```
##
##   0   1
## 102 399
```

```
table(data$respondent_t3)
```

```
##
##   0   1
## 100 401
```

```
table(data$respondent_t4)
```

```
##
##   0   1
## 116 385
```

- Although not described in the main text, there is a clear effect on the index of all primary outcomes pre-specified in advance.

```
est.ate(data$all.dvs.t1)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.225"    "0.072"    "3.112"   "0.001*"
```

- "Before the intervention, the treatment and placebo groups scored similarly on this index." See also tables S13-S17.

```
est.ate(data$trans.tolerance.dv.t0, include.covariates = FALSE)
```

```
##   Estimate Std. Error    t value        *p*
##   "-0.013"    "0.137"   "-0.098"   "*n.s.*"
```

- "After the intervention, the treatment group was considerably more accepting of transgender people than the placebo group ($t = 4.03$; $p < 0.001$)."

```
est.ate(data$trans.tolerance.dv.t1)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.292"    "0.072"    "4.028"   "0.000*"
```

- "These effects are substantively large: these brief conversations increased positivity towards transgender people on a feeling thermometer (20, 21) by approximately 10 points, an amount larger than the secular increase in positive affect towards gay men and lesbians among Americans between 1998 and 2012 (8.5 points; see table S22)."

```
est.ate(data$therm_trans_t1)
```

```
##   Estimate Std. Error    t value        *p*
##   "10.396"    "2.433"    "4.273"   "0.000*"
```

To contextualize these results, we compare the treatment effect observed in the current study to the change over time in American public opinion towards gay men using the feeling thermometer in the American National Election Studies (*46*). The treatment effect in the current study on the transgender feeling thermometer at 3 days was 10.396, which is larger than the change in the ANES feeling thermometer for gay men from 1998 to 2012 (see Table S22).

- "Fig. 1 shows the point estimates at each wave and reveals that these effects persisted longitudinally: the treatment group remained more accepting in every follow-up survey (t-tests; all ps < 0.01)."

Replication code for Fig 1. is available below. Below are the estimates for each follow-up wave. Note that the point estimates are not necessarily comparable across waves because the composition of respondents changes slightly and the factor-analysis is re-computed at each wave.

```
est.ate(data$trans.tolerance.dv.t2)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.238"    "0.079"     "3.03"   "0.001*"
```

```
est.ate(data$trans.tolerance.dv.t3)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.352"    "0.074"    "4.736"   "0.000*"
```

```
est.ate(data$trans.tolerance.dv.t4)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.345"    "0.078"    "4.441"   "0.000*"
```

- "The intervention was also broadly effective: the effects are significant at the $p < 0.01$ level for both registered Democrats and registered Republicans (see table S6) and for those who began more and less supportive than average (t-tests; see table S8)."

Please refer to the section entitled "Heterogenous Treatment Effects."

- "Finally, conversations with transgender and non-transgender canvassers were both effective (t-tests; see table S5). Tables S1-S4 show estimates for each item at each wave."

Please see the section entitled "Comparison of Transgender and Non-Transgender canvassers."

```
est.ate(data$trans.tolerance.dv.t1, data$canvasser_trans == 1)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.453"    "0.109"    "4.145"   "0.000*"
```

```
est.ate(data$trans.tolerance.dv.t1, data$canvasser_trans == 0)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.191"    "0.094"    "2.027"   "0.022*"
```

```
est.ate(data$all.dvs.t1, data$canvasser_trans == 1)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.277"    "0.103"    "2.676"   "0.004*"
```

```
est.ate(data$all.dvs.t1, data$canvasser_trans == 0)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.172"    "0.098"    "1.747"   "*n.s.*"
```

- "there was no difference between the treatment and placebo groups' support for a law protecting transgender people from discrimination in the three-day and three-week follow-up surveys."

```
est.ate(data$miami_trans_law_t1_avg)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.002"    "0.135"    "0.016"   "*n.s.*"
```

```
est.ate(data$miami_trans_law_t2_avg)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.041"    "0.141"    "0.293"   "*n.s.*"
```

- "treatment subjects were 0.36 scale points more supportive of the law protecting this group than treatment subjects once the survey question defined the term transgender ($t = 2.20$; $p < 0.05$)."

```
est.ate(data$miami_trans_law_t3_avg)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.362"    "0.164"    "2.203"   "0.014*"
```

- "First, the intervention's effect withstood this attack: the canvassing intervention treatment group remained 0.40 scale points more supportive of the law than the placebo group ($t = 1.77$; $p < 0.05$)."

```
est.ate(data$trans_law_post_ad_t3)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.401"    "0.227"    "1.768"   "0.039*"
```

- "However, subjects in the canvassing treatment group remained 0.30 scale points more supportive than placebo subjects in the three-month survey ($t = 1.94$; $p < 0.05$)."

```
est.ate(data$miami_trans_law_t4_avg)
```

```
##   Estimate Std. Error    t value        *p*
##    "0.295"    "0.152"    "1.937"   "0.027*"
```

- "The intervention was effective among all pre-specified subgroups, including political parties."

Please refer to the section entitled "Heterogenous Treatment Effects."

- "Canvassers did not require extensive experience: both first-time and experienced canvassers were effective."

```
est.ate(data$trans.tolerance.dv.t1, data$canvasser_experience=="No")
```

```
##   Estimate Std. Error    t value        *p*
##    "0.249"    "0.076"    "3.277"   "0.001*"
```

```
est.ate(data$trans.tolerance.dv.t1, data$canvasser_experience=="Yes")
```

```
##   Estimate Std. Error    t value        *p*
##    "0.702"    "0.346"     "2.03"   "0.024*"
```

Code to Produce Figures

Code to Produce Figure 1 The below code was used to produce Figure 1.

```
# Estimate the treatment effect for ALL canvassers
t1.all <- est.ate(data$trans.tolerance.dv.t1)
t2.all <- est.ate(data$trans.tolerance.dv.t2)
t3.all <- est.ate(data$trans.tolerance.dv.t3)
t4.all <- est.ate(data$trans.tolerance.dv.t4)

# Estimate the treatment effect by trans and cis canvassers
#At 3 day survey
t1.trans <- est.ate(data$trans.tolerance.dv.t1,
                    data$canvasser_trans == 1)
t1.cis <- est.ate(data$trans.tolerance.dv.t1,
                  data$canvasser_trans == 0)

#At 3 week survey
t2.trans <- est.ate(data$trans.tolerance.dv.t2,
                    data$canvasser_trans == 1)
t2.cis <- est.ate(data$trans.tolerance.dv.t2,
                  data$canvasser_trans == 0)

#At 6 week survey
t3.trans <- est.ate(data$trans.tolerance.dv.t3,
                    data$canvasser_trans == 1)
t3.cis <- est.ate(data$trans.tolerance.dv.t3,
                  data$canvasser_trans == 0)
```

```
#At 3 month survey
t4.trans <- est.ate(data$trans.tolerance.dv.t4,
                    data$canvasser_trans == 1)
t4.cis <- est.ate(data$trans.tolerance.dv.t4,
                  data$canvasser_trans == 0)


# Make DF of summary stats
summary.stats.df <- as.data.frame(rbind(t1.all, t2.all,
                          t3.all, t4.all,
                          t1.trans, t1.cis,
                          t2.trans, t2.cis,
                          t3.trans, t3.cis,
                          t4.trans, t4.cis),
                          stringsAsFactors = FALSE)

# Change from strings back to numeric, and remove t- and p-values, which are not used.
summary.stats.df <- summary.stats.df[,1:2]
summary.stats.df[,1] <- as.numeric(summary.stats.df[,1])
summary.stats.df[,2] <- as.numeric(summary.stats.df[,2])

# Better variable names
names(summary.stats.df) <- c("point.estimate", "se")

# Map row names of summary.stats.df into days
unique.days <- c(3, 3*7, 6*7, 12*7)
summary.stats.df$days <- unique.days[as.numeric(substr(row.names(summary.stats.df), 2, 2))]

# Read canvasser group from row names
canvasser.label.map <- list(all = "All",
                            tra = "Transgender/Gender\nNon-Conforming Only",
                            cis = "Non-Transgender Only")
summary.stats.df$Canvasser <- factor(as.character(
  canvasser.label.map[substr(row.names(summary.stats.df), 4, 6)]
  ))

# X position of different canvasser groups
summary.stats.df$xpos <- with(summary.stats.df, days + as.numeric(Canvasser) -
                                mean(as.numeric(Canvasser)))

# Point estimate Y
summary.stats.df$point.estimate.y <- summary.stats.df$point.estimate
  # Fix text overlap of point estimate labels.
  summary.stats.df$point.estimate.y[9] <- summary.stats.df$point.estimate.y[9] + .003
  summary.stats.df$point.estimate.y[3] <- summary.stats.df$point.estimate.y[3] - .003

# Compute CIs
summary.stats.df$se.high <- summary.stats.df$point.estimate + summary.stats.df$se
summary.stats.df$se.low <- summary.stats.df$point.estimate - summary.stats.df$se
summary.stats.df$ci.high <- summary.stats.df$point.estimate + summary.stats.df$se * 1.96
summary.stats.df$ci.low <- summary.stats.df$point.estimate - summary.stats.df$se * 1.96

summary.stats.df$point.estimate.label <- paste0(round(summary.stats.df$point.estimate,
```

```
                                                    2)," SDs")

g <- ggplot(summary.stats.df,
            aes(x=xpos, y=point.estimate,
                group=Canvasser, color=Canvasser)) +
  theme_classic() +
  # CIs
  geom_linerange(aes(ymin=se.low, ymax=se.high), lwd=1) +
  geom_linerange(aes(ymin=ci.low, ymax=ci.high)) +
  # Point estimate points
  geom_point(color="black") +
  # Point estimate markers
  annotate("text", label=summary.stats.df$point.estimate.label,
           x = summary.stats.df$xpos + 6.3,
           y = summary.stats.df$point.estimate.y,
           size = 3) +
  # Canvassing treatment line / label
  geom_vline(xintercept = 0, linetype = "dashed") +
  annotate("text", label = "Canvassing Treatment",
           x = -1.5, y = .3, size = 3.5, angle = 90) +
  # Day labels
  annotate("text",
           label = c("+3 Days", "+3 Weeks", "+6 Weeks", "+3 Months"),
           x = unique.days, y = -.04,
           colour = "black", size = 3) +
  # Y axis
  ylab("Effect on Transgender Tolerance Scale, in Standard Deviations") +
  scale_y_continuous() +
  # X axis
  xlab("Days After Canvassing Treatment") +
  geom_hline(yintercept = 0) +
  # Overall Title and Legend
  ggtitle("Differences Between Treatment and Placebo") +
  guides(fill=FALSE) +
  theme(legend.position = "bottom")
ggsave("figure1.pdf", g, width=8, height=6, units="in")
```

Code to Produce Figure 2 The below code was used to produce Figure 2.

```
# Multiple plot function (from http://www.cookbook-r.com/Graphs/
#                             Multiple_graphs_on_one_page_(ggplot2)/)
#
# ggplot objects can be passed in ..., or to plotlist (as a list of ggplot objects)
# - cols:   Number of columns in layout
# - layout: A matrix specifying the layout. If present, 'cols' is ignored.
#
# If the layout is something like matrix(c(1,2,3,3), nrow=2, byrow=TRUE),
# then plot 1 will go in the upper left, 2 will go in the upper right, and
# 3 will go all the way across the bottom.
#
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)
```

```
  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

 if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
  }
}
```

The STATA code below creates covariate-adjusted clustered standard errors for the data points, outputting them in fig2margins.csv, which we then read in below to display as a graph. (We were unable to locate a function in R that did this.)

```
clear all
cd "~/Google Drive/Broockman-Kalla/Trans Prejudice in Miami/Replication File/"
use broockman_kalla_replication_data.dta, clear

//////////////////
*Housekeeping
keep if contacted==1
sum vf_age
replace vf_age =r(mean) if missing(vf_age)
assert !missing(vf_age)
replace survey_language_es = 1 if survey_language_t0 == "ES" & missing(survey_language_es)
replace survey_language_es = 0 if survey_language_t0 == "EN" & missing(survey_language_es)
assert !missing(survey_language_es)
//////////////////

//////////////////
```

```
*Law DV.
*Create outcome scale by averaging over the two questions.
gen miami_trans_law_t0_avg = (miami_trans_law_t0 + miami_trans_law2_t0)/2
gen miami_trans_law_t1_avg = (miami_trans_law_t1 + miami_trans_law2_t1)/2
gen miami_trans_law_t2_avg = (miami_trans_law_t2 + miami_trans_law2_t2)/2
*Note: Beginning with t3, the definition was added.
gen miami_trans_law_t3_avg = (miami_trans_law_withdef_t3 + miami_trans_law2_withdef_t3)/2
*Note: Only one question was asked in t3 after the ad was shown, so no averaging is required.
gen miami_trans_law_t4_avg = (miami_trans_law_withdef_t4 + miami_trans_law2_withdef_t4)/2
//////////////////

*Covariates, per pre-analysis plan
local covars miami_trans_law_t0 miami_trans_law2_t0 therm_trans_t0 ///
gender_norms_sexchange_t0 gender_norms_moral_t0 gender_norms_abnormal_t0 ///
ssm_t0 therm_obama_t0 therm_gay_t0 vf_democrat ideology_t0 ///
religious_t0 exposure_gay_t0 exposure_trans_t0 pid_t0 sdo_scale ///
gender_norm_daugher_t0 gender_norm_looks_t0 ///
gender_norm_rights_t0 therm_afams_t0 vf_female vf_hispanic ///
vf_black vf_age survey_language_es cluster_level_t0_scale_mean ///

reg miami_trans_law_t1_avg i.treat_ind `covars', cluster(hh_id)
estadd margins i.treat_ind
estout using fig2margins.csv, cells("margins_b margins_se") replace delimiter(",")

reg miami_trans_law_t2_avg i.treat_ind `covars', cluster(hh_id)
estadd margins i.treat_ind
estout using fig2margins.csv, cells("margins_b margins_se") append delimiter(",")

reg miami_trans_law_t3_avg i.treat_ind `covars', cluster(hh_id)
estadd margins i.treat_ind
estout using fig2margins.csv, cells("margins_b margins_se") append delimiter(",")

reg trans_law_post_ad_t3 i.treat_ind `covars', cluster(hh_id)
estadd margins i.treat_ind
estout using fig2margins.csv, cells("margins_b margins_se") append delimiter(",")

reg miami_trans_law_t4_avg i.treat_ind `covars', cluster(hh_id)
estadd margins i.treat_ind
estout using fig2margins.csv, cells("margins_b margins_se") append delimiter(",")

//Clean up data
insheet using fig2margins.csv, clear
rename v1 lawcondition
    replace lawcondition = "Placebo Group Mean" if lawcondition == "0.treat_ind"
    replace lawcondition = "Treatment Group Mean" if lawcondition == "1.treat_ind"
drop if v2 == "."  | v2 == "margins_b"
rename v2 lawmean
rename v3 lawse
outsheet using fig2margins.csv, comma replace
```

```
days <- c(3, # 3 day survey
          21, # 3 week survey
          39,45, # 6 week survey; for presentational purposes, two estimates offset +/- 3 days
          84) # 3 month survey
```

```
#PANEL A
#Make DF of summary stats
#Read in Stata output
summary.stats.law.df <- read.csv(paste0(wd, 'fig2margins.csv'))
summary.stats.law.df$x.position <- c(
  rep(days[1] + .45,2), rep(days[2],2), rep(days[3],2),
      rep(days[4],2), rep(days[5]-.45,2)) +
  rep(c(-0.45,.45),5) # Jitter
summary.stats.law.df$law.color.codes <- rep(c("gray40", "#F8766D"),5)

summary.stats.law.df$law.mean <- as.numeric(summary.stats.law.df$lawmean)
summary.stats.law.df$law.se <- as.numeric(summary.stats.law.df$lawse)
summary.stats.law.df$x.position <- as.numeric(summary.stats.law.df$x.position)

# Compute CIs
summary.stats.law.df$se.high <- summary.stats.law.df$law.mean + summary.stats.law.df$law.se
summary.stats.law.df$se.low <- summary.stats.law.df$law.mean - summary.stats.law.df$law.se
summary.stats.law.df$ci.high <- summary.stats.law.df$law.mean + summary.stats.law.df$law.se*1.96
summary.stats.law.df$ci.low <- summary.stats.law.df$law.mean - summary.stats.law.df$law.se*1.96

fig2.panela <- ggplot(data=summary.stats.law.df,
                          aes(x=x.position, y=law.mean, color=lawcondition)) +
  geom_point() +
  scale_colour_manual(breaks = summary.stats.law.df$lawcondition,
                      values = unique(as.character(summary.stats.law.df$law.color.codes))) +
  # Theme
  theme_classic() +
  # CIs
  geom_linerange(aes(x=x.position, ymin=se.low, ymax=se.high), lwd=1) +
  geom_linerange(aes(x=x.position, ymin=ci.low, ymax=ci.high)) +
  # Legend
  theme(legend.position=c(.725,.1),
        legend.title = element_blank(),
        legend.background = element_rect(color="black"),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank()) +
  # Axes
  ylab("Mean Support for\nNon-Discrimination Law") +
  xlab("Days After Canvassing Treatment") +
  scale_y_continuous(breaks = c(-.3,.2,.7,1.2,1.7)) +
  # change axis ticks to include a - sign so numbers line up
  # Canvassing treatment line
  geom_vline(xintercept = 0, linetype = "dashed") +
  annotate("text", label = "Canvassing Treatment",
           x = 1.5, y = 1.075, ymin = 0, size = 4, angle = 90) +
  # Define trans line
  geom_vline(xintercept = 36, linetype = "dashed", colour = "darkgreen") +
  annotate("text", label = "Define Transgender",
           x = 37, y = 1.075, ymin = 0, size = 4, angle = 90, colour = "darkgreen") +
  # Show video line
  geom_vline(xintercept = 42, linetype = "dashed", colour = "purple") +
  annotate("text", label = "Show Opposition Video",
           x = 43, y = 1.075, ymin = 0, size = 4, angle = 90, colour = "purple") +
```

```
  # Day labels
  annotate("text",
           label = c("+3 Days", "+3 Weeks", "+6 Weeks", "+3 Months"),
           x = c(3, 21, 42, 84), y = -0.23,
           colour = "black", size = 4) +
  ggtitle("Panel A. Support for Non-Discrimination Law\n(Condition Means)")

#PANEL B.
#Now, plot the treatment effect.
# Estimate the treatment effect for ALL canvassers
t1.law.effect <- est.ate(data$miami_trans_law_t1_avg)
t2.law.effect <- est.ate(data$miami_trans_law_t2_avg)
t3.law.effect <- est.ate(data$miami_trans_law_t3_avg)
t3.law.effect.post.ad <- est.ate(data$trans_law_post_ad_t3)
t4.law.effect <- est.ate(data$miami_trans_law_t4_avg)

# Make DF of summary stats
summary.stats.law.effect.df <- as.data.frame(rbind(t1.law.effect, t2.law.effect,
                                        t3.law.effect, t3.law.effect.post.ad,
                                        t4.law.effect),
                          stringsAsFactors = FALSE)

# Change from strings back to numeric, and remove t- and p-values, which are not used.
summary.stats.law.effect.df <- summary.stats.law.effect.df[,1:2]
summary.stats.law.effect.df[,1] <- as.numeric(summary.stats.law.effect.df[,1])
summary.stats.law.effect.df[,2] <- as.numeric(summary.stats.law.effect.df[,2])

# Better variable names
names(summary.stats.law.effect.df) <- c("point.estimate", "se")

# Read days
summary.stats.law.effect.df$x.position <- days

# Compute CIs
summary.stats.law.effect.df$se.high <- summary.stats.law.effect.df$point.estimate +
        summary.stats.law.effect.df$se
summary.stats.law.effect.df$se.low <- summary.stats.law.effect.df$point.estimate -
        summary.stats.law.effect.df$se
summary.stats.law.effect.df$ci.high <- summary.stats.law.effect.df$point.estimate +
        summary.stats.law.effect.df$se * 1.96
summary.stats.law.effect.df$ci.low <- summary.stats.law.effect.df$point.estimate -
        summary.stats.law.effect.df$se * 1.96
summary.stats.law.effect.df$label <- factor("Treatment Effect")

fig2.panelb <- ggplot(summary.stats.law.effect.df,
            aes(x=x.position, y=point.estimate, color=label)) +
  theme_classic() +
  geom_point() +
  # CIs
  geom_linerange(aes(ymin=se.low, ymax=se.high), lwd=1) +
  geom_linerange(aes(ymin=ci.low, ymax=ci.high)) +
  # Legend
  theme(legend.position=c(.725,.1),
```

```
        legend.title = element_blank(),
        legend.background = element_rect(color="black"),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank()) +
  # Axes
  ylab("Treatment Effect of Canvassing on\nSupport For Non-Discrimination Law") +
  xlab("Days After Canvassing Treatment") +
  geom_hline(yintercept = 0, linetype = "dashed") +
  # "Canvassing treatment"" line
  geom_vline(xintercept = 0, linetype = "dashed") +
  annotate("text", label = "Canvassing Treatment",
           x = 1.5, y = 0.5, ymin = 0, size = 4, angle = 90) +
  # "Define transgender"" line
  geom_vline(xintercept = 36, linetype = "dashed", colour = "darkgreen") +
  annotate("text", label = "Define Transgender",
           x = 37, y = 0.5, ymin = 0, size = 4, angle = 90, colour = "darkgreen") +
  # "Show opposition video"" line
  geom_vline(xintercept = 42, linetype = "dashed", colour = "purple") +
  annotate("text", label = "Show Opposition Video",
           x = 43, y = 0.5, ymin = 0, size = 4, angle = 90, colour = "purple") +
  # Day labels
  annotate("text",
           label = c("+3 Days", "+3 Weeks", "+6 Weeks", "+3 Months"),
           x = c(3, 21, 42, 84), y = -0.3,
           color = "black", size = 4) +
  ggtitle("Panel B. Treatment Effect at Each Wave\n(Differences Between The Condition Means Shown Above)

pdf("figure2.pdf", width = 8.5, height = 9)
xl <- xlim(0,87)
multiplot(fig2.panela + xl, fig2.panelb + xl)
dev.off()
```

```
## pdf
##   2
```

Intervention Details

Training

Before each canvass, the Los Angeles LGBT Center and SAVE spent 2.5 hours training volunteers to improve their ability to effectively deliver the intervention. The trainings focused on providing volunteers with the skills to listen to voters, focus on soliciting voters' experiences, and asking them follow-up questions that would encourage perspective-taking. For example, volunteer canvassers would role play and view video of past canvass conversations.

Intervention Procedure

The canvassers were trained to follow the below procedure when approaching homes when subjects were in the treatment condition. 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 in points 6 and 7, soliciting a personal experience from voters and asking them to consider transgender people's perspective.

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.

Encourage Active Processing

2. **Intervention begins: inform subject they will soon face a decision about whether to protect transgender people from discrimination.** The canvassers began the intervention by engaging in a series of strategies that have been shown to facilitate active processing of persuasive messages. First, canvassers informed voters they might face a decision about whether to protect transgender people from discrimination (the potential referendum election about whether to repeal the law). Informing people that they may face a decision about a topic encourages active processing of information about it (*47*).
3. **Solicit subject opinion about the law and ask subject to explain it.** Canvassers asked voters about their opinion on the law – do they support or oppose it, or are they not sure? – 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 further effortful reflection and rapport.
4. **Show video with image of transgender person and both sides of the argument.** Canvassers next showed voters a video of a local news clip discussing the law that depicted advocates on both sides of the issue articulating their arguments (the video is available at https://www.youtube.com/watch?v=XZxARafQrZY) and then asked the voters to talk aloud about their reactions to the video. Showing this video was expected to produce feelings of ambivalence and uncertainty, which also facilitates active processing (*48*). The argument on the supportive side was that all people should be treated equally; the argument on the opposing side echoed a common refrain, that transgender women are simply men pretending to be women who want to enter women's bathrooms. The video also depicted a transgender person and defined the term. This helped subjects who were unfamiliar with the term transgender understand what the term meant. Further, showing an image of a member of an outgroup is one common entry point for perspective-taking (*26*). If the canvassers were transgender, they would reveal that they were transgender by this point as well.
5. **Ask subject how they reacted to the video.** To further build rapport and encourage active processing, the subjects were asked to expound upon their reaction to the video and any uncertainty it induced.

Encourage Perspective-Taking

6. **Solicit a personal experience with judgment.** The intervention attempted to prompt "analogic perspective-taking" (*16*) by first asking voters to describe if they had ever felt the kind of negative judgment or stigma that the transgender person in the video felt. If necessary, canvassers sometimes told their own stories of judgment in order to make voters feel comfortable sharing a story of their own. At this point, transgender canvassers would often tell a personal story of facing judgment and discrimination as transgender.
7. **Encourage analogic perspective-taking: subject's personal experiences of judgment provide a way to take transgender people's perspective.** After voters told a story, canvassers

encouraged subjects to think about how their personal stories of judgment for being different provided a way for them to understand what it would be like to be transgender and understand transgender people's experiences of judgment, including experiencing the kinds of negative judgment the subject themselves might have cast earlier in the conversation. For example, one of the canvassers reported that one voter who began unsympathetic was a military veteran who told a story about being denied for jobs because he suffered from post-traumatic stress disorder. This voter relayed his impression that potential employers generalized about his character from this one fact about him. This voter then reported thinking about what it would be like for a transgender person if an employer did not hire them because of their identity and realized that he and such a transgender person would be similarly situated.

Encourage Active Processing

8. **Ask for opinion again; rehearse opinion change.** The intervention ended with a final attempt to encourage active processing of the implications of taking transgender people's perspective, with canvassers asking voters if and why the conversation changed their attitudes towards the group or the law protecting them. Rehearsal of opinion change is another strategy that has been shown to facilitate active processing and increase the persistence of attitude changes (*49*). 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 told voters that they may vote on an initiative that would require supermarkets to charge for plastic bags instead of giving them away for free. The canvasser asked voters how they felt about this law.
3. **Conversation ends.** The canvasser thanks the subject and leaves.

Opposition Ad Treatments for Figure 2

As described in the text, we showed subjects an opposition video ad near the end of the six-week survey, after they had responded to miami_trans_law_withdef_t3 and miami_trans_law2_withdef_t3.

In an updated pre-analysis plan filed prior to analyzing the third post-treatment survey, we noted, "Because we did not see any effect on the miami_trans_law items, we removed these items. However, one possibility is that people in the control group indicated they were supportive because they did not know what transgender meant. We therefore reformulated the law items to include a definition of transgender."

We further noted, "We will also look to see if there is a difference on people's views on the law after having seen a transphobic advertisement from a previous election campaign." We did not realize we were able to embed ads in the online survey until this point in time, hence the update to the pre-analysis plan.

Marijuana Ad First

To limit suspicion, we first showed subjects an ad from a campaign about marijuana legalization (marijuana attitudes were another topic on the surveys): "With election season coming up there's lots of ads about various issues and candidates. We want to show you some ads from other places about issues that are being discussed in Miami. First, we want to hear what you think about the ad below, which is from Alaska. Please be sure to turn your sound on and be sure the volume is high enough. Once you are ready, click on the ad to play it."

The following YouTube video was then shown on their screens:

- Marijuana Ad https://www.youtube.com/watch?v=OEVRpeFGaU4

Early in the survey, respondents were first asked: ''Earlier this month, Miami-Dade County and the city of Miami Beach changed the laws to 'decriminalize' marijuana, meaning that when police catch people with only a small amount of marijuana, they are allowed to issue them small fines instead of arrest them. Supporters of this new law say it will save the city money and allow the police to focus on more serious crimes. Opponents say marijuana is a 'gateway' drug that is important to control. What do you think? When it comes to regulating marijuana use, what do you think is right for Miami-Dade?" 24.6% of treatment group respondents and 21.9% of placebo group respondents stated that "Miami-Dade should allow people to use marijuana for any reason (e.g., for recreational purposes)". This difference between treatment and placebo opinions on marijuana is not statistically significant (p=0.52).

Subjects then viewed the above marijuana ad, which advocated that "The war on marijuana is wasteful" and that marijuana should be regulated like alcohol. After viewing the ad, subjects were then asked: "As we've mentioned, Florida is considering changing the laws that prohibit the use of marijuana. After seeing the ad, when it comes to allowing use of marijuana, what do you think is right for Florida?" At this point, 37.2% of treatment group respondents and 33.8% of placebo group respondents stated that "Miami-Dade should allow people to use marijuana for any reason (e.g., for recreational purposes)". This difference between treatment and placebo opinions on marijuana is not statistically significant (p=0.48).

We did not ask about marijuana in the 3-month survey.

Transphobic Opposition Ads

We next showed subjects one of the three opposition ads below. Subjects were randomly assigned to these three ads with equal probability. We showed subjects one of three ads instead of picking one ad in order to increase the external validity of this manipulation; real opposition campaigns employ all the approaches below.

- Opposition Ad from Olympia, Washington: https://www.youtube.com/watch?v=gmjztjHe15s
- Opposition Ad from Anchorage, Alaska: https://www.youtube.com/watch?v=o8yoAaVgJVo
- Opposition Ad Kalamazoo, Michigan: https://www.youtube.com/watch?v=KfBLSw2OHro

We asked "Did the video play correctly for you?" and over 91% of subjects said it did.

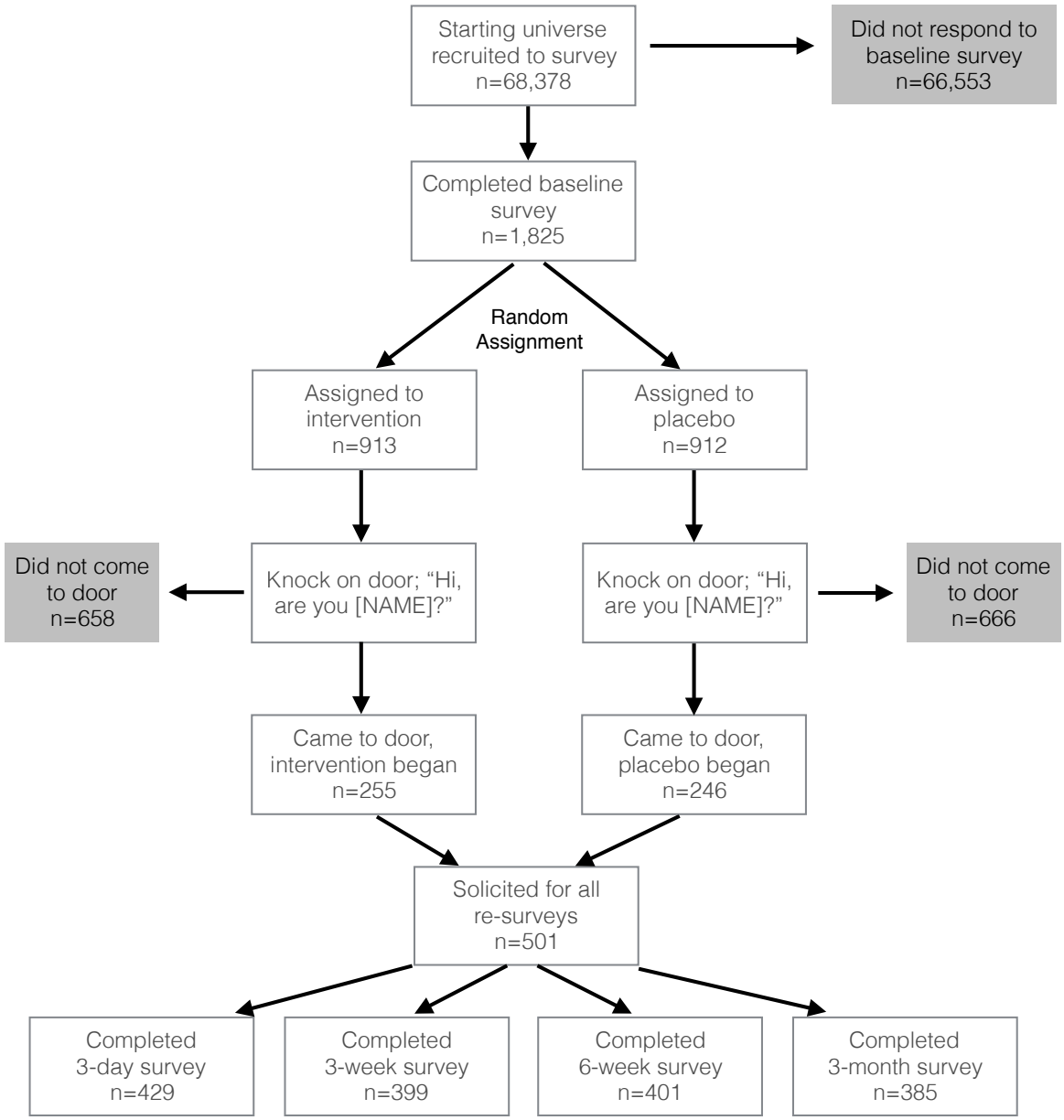
On the screen following the video, we asked trans_law_post_ad_t3. Recall that Figure 2 presents the results of this manipulation.

Variable Details and Codebook

Table S23 provides the mean and standard deviation for all numeric variables by treatment condition. Table S24 provides a codebook that describes all of the variables included in the replication file.

Fig. S1.

Fig. S1.



Overview of experimental design.

Table S1 to S24.

Table S1.

	Estimate	Std. Error	t value	p
all.dvs.t1	0.225	0.072	3.112	0.001*
trans.tolerance.dv.t1	0.292	0.072	4.028	0.000*
miami_trans_law_t1_avg	0.002	0.135	0.016	*n.s.*
miami_trans_law_t1	-0.084	0.154	-0.547	*n.s.*
miami_trans_law2_t1	0.089	0.144	0.615	*n.s.*
therm_trans_t1	10.396	2.433	4.273	0.000*
gender_norm_sexchange_t1	0.151	0.12	1.263	*n.s.*
gender_norm_moral_t1	0.198	0.104	1.912	0.028*
gender_norm_abnormal_t1	0.307	0.114	2.696	0.004*
gender_norm_trans_moral_wrong_t1	0.253	0.109	2.333	0.01*
gender_nonconformity_t1	0.148	0.088	1.684	*n.s.*

Results by Individual Item in the Three Day Survey. We present the covariate-adjusted complier average causal effect estimated using the procedure described above and that was pre-specified in advance. The first row is the treatment effect on the first factor produced from the pre-specified index of all primary outcomes, the next row is the treatment effect on the the first factor produced from the transgender prejudice scale, the next row is the treatment effect on the average of the two items about the law, and the remaining rows are the treatment effects on each of the individual items that comprise the scales. All standard errors are cluster-robust standard errors, clustering on household to reflect the random assignment.

Table S2.

	Estimate	Std. Error	t value	p
all.dvs.t2	0.185	0.076	2.423	0.008*
trans.tolerance.dv.t2	0.238	0.079	3.03	0.001*
miami_trans_law_t2_avg	0.041	0.141	0.293	*n.s.*
miami_trans_law_t2	-0.113	0.17	-0.662	*n.s.*
miami_trans_law2_t2	0.195	0.146	1.339	*n.s.*
therm_trans_t2	7.447	2.693	2.765	0.003*
gender_norm_sexchange_t2	0.089	0.136	0.649	*n.s.*
gender_norm_moral_t2	0.208	0.114	1.835	0.034*
gender_norm_abnormal_t2	0.349	0.116	3.011	0.001*
gender_norm_trans_moral_wrong_t2	0.185	0.118	1.561	*n.s.*
trans_teacher_t2	0.217	0.138	1.57	*n.s.*
trans_bathroom_t2	0.215	0.137	1.573	*n.s.*
gender_nonconformity_t2	0.2	0.083	2.412	0.008*

Results by Individual Item in the Three Week Survey. We present the covariate-adjusted complier average causal effect estimated using the procedure described above and that was pre-specified in advance. The first row is the treatment effect on the first factor produced from the pre-specified index of all primary outcomes, the next row is the treatment effect on the the first factor produced from the transgender prejudice scale, the next row is the treatment effect on the average of the two items about the law, and the remaining rows are the treatment effects on each of the individual items that comprise the scales. All standard errors are cluster-robust standard errors, clustering on household to reflect the random assignment.

Table S3.

	Estimate	Std. Error	t value	p
all.dvs.t3	0.295	0.076	3.874	0.000*
trans.tolerance.dv.t3	0.352	0.074	4.736	0.000*
miami_trans_law_t3_avg	0.362	0.164	2.203	0.014*
miami_trans_law_withdef_t3	0.276	0.181	1.519	*n.s.*
miami_trans_law2_withdef_t3	0.448	0.174	2.576	0.005*
therm_trans_t3	8.323	2.523	3.298	0.001*
gender_norm_sexchange_t3	0.354	0.132	2.688	0.004*
gender_norm_moral_t3	0.254	0.109	2.325	0.01*
gender_norm_abnormal_t3	0.359	0.112	3.215	0.001*
gender_norm_trans_moral_wrong_t3	0.228	0.11	2.063	0.02*
trans_teacher_t3	0.415	0.143	2.911	0.002*
trans_bathroom_t3	0.446	0.131	3.392	0.000*
gender_nonconformity_t3	0.171	0.078	2.189	0.015*

Results by Individual Item in the Six Week Survey. We present the covariate-adjusted complier average causal effect estimated using the procedure described above and that was pre-specified in advance. The first row is the treatment effect on the first factor produced from the pre-specified index of all primary outcomes, the next row is the treatment effect on the the first factor produced from the transgender prejudice scale, the next row is the treatment effect on the average of the two items about the law, and the remaining rows are the treatment effects on each of the individual items that comprise the scales. All standard errors are cluster-robust standard errors, clustering on household to reflect the random assignment.

Table S4.

	Estimate	Std. Error	t value	p
all.dvs.t4	0.257	0.075	3.429	0.000*
trans.tolerance.dv.t4	0.345	0.078	4.441	0.000*
miami_trans_law_t4_avg	0.295	0.152	1.937	0.027*
miami_trans_law_withdef_t4	0.286	0.176	1.627	*n.s.*
miami_trans_law2_withdef_t4	0.303	0.16	1.889	0.03*
therm_trans_t4	9.224	2.676	3.447	0.000*
gender_norm_sexchange_t4	0.207	0.116	1.789	0.037*
gender_norm_moral_t4	0.194	0.117	1.657	*n.s.*
gender_norm_abnormal_t4	0.24	0.125	1.925	0.028*
gender_norm_trans_moral_wrong_t4	0.318	0.117	2.722	0.003*
trans_teacher_t4	0.385	0.135	2.848	0.002*
trans_bathroom_t4	0.59	0.132	4.481	0.000*
gender_nonconformity_t4	0.417	0.088	4.722	0.000*

Results by Individual Item in the Three Month Survey. We present the covariate-adjusted complier average causal effect estimated using the procedure described above and that was pre-specified in advance. The first row is the treatment effect on the first factor produced from the pre-specified index of all primary outcomes, the next row is the treatment effect on the the first factor produced from the transgender prejudice scale, the next row is the treatment effect on the average of the two items about the law, and the remaining rows are the treatment effects on each of the individual items that comprise the scales. All standard errors are cluster-robust standard errors, clustering on household to reflect the random assignment.

Table S5.

	Estimate	Std. Error	t value	p
Trans T1	0.453	0.109	4.145	0.000*
Non-Trans T1	0.191	0.094	2.027	0.022*
Trans T2	0.3	0.128	2.334	0.011*
Non-Trans T2	0.191	0.098	1.954	0.026*
Trans T3	0.364	0.114	3.193	0.001*
Non-Trans T3	0.321	0.093	3.442	0.000*
Trans T4	0.427	0.135	3.169	0.001*
Non-Trans T4	0.324	0.104	3.105	0.001*

Separately Estimating Effects of Transgender and Non-Transgender Canvassers. We separately compute the treatment effects on the trans.tolerance.dvs.t# factor for voters who were canvassed by transgender canvassers and by non-transgender canvassers at each survey wave. We find that both types of canvassers are effective at reducing negative attitudes towards transgender people. Because of the turf-level random assignment of canvassers the standard errors presented in this table (which pertain only to treatment effects among the subpopulations) do not allow for the test of the hypothesis that transgender canvassers had larger treatment effects than non-transgender canvassers.

Table S6.

	Estimate	Std. Error	t value	p
Democrats	0.276	0.106	2.608	0.005*
Republicans	0.45	0.143	3.144	0.001*
Independents	0.138	0.162	0.852	*n.s.*

Effects at 3 Days by Party Registration, separately. Linear regression with clustered standard errors.

Table S7.

	Estimate	Std. Error	t	p
Intercept	-0.12	0.066	-1.9	0.061
Treatment	0.33	0.1	3.3	0.0011
Democrat	0.0057	0.084	0.067	0.95
Independent	0.087	0.097	0.9	0.37
Democrat x Treat	-0.096	0.12	-0.77	0.44
Independent X Treat	-0.3	0.14	-2	0.042

Effects at 3 Days by Party Registration, Interaction Term. Linear regression with clustered standard errors.

Table S8.

	Estimate	Std. Error	t value	p
Low Baseline	0.331	0.134	2.464	0.007*
High Baseline	0.236	0.089	2.652	0.004*

Effects at 3 Days by Baseline, separately. Linear regression with clustered standard errors.

Table S9.

	Estimate	Std. Error	t	p
Intercept	-0.098	0.036	-2.7	0.0065
Treat	0.21	0.052	4	8.3e-05
Democrat	0.018	0.04	0.44	0.66
Baseline X Treat	-0.029	0.055	-0.52	0.6

Effects at 3 Days by Baseline, Interaction Term. Linear regression with clustered standard errors.

Table S10.

	Estimate	Std. Error	t value	p
Low NFC	0.344	0.136	2.525	0.006*
High NFC	0.217	0.094	2.3	0.011*

Effects at 3 Days by Need for Cognition, separately. Linear regression with clustered standard errors.

Table S11.

	Estimate	Std. Error	t	p
Intercept	-0.087	0.038	-2.3	0.022
Treat	0.2	0.054	3.7	0.00023
NFC	0.054	0.046	1.2	0.23
NFCxTreat	-0.087	0.066	-1.3	0.19

Effects at 3 Days by Need for Cognition, Interaction Term. Linear regression with clustered standard errors.

Table S12.

	Placebo	Treatment	t-statistic
Age	48	46	-1.66
Female	0.58	0.59	0.45
Black	0.25	0.25	-0.01
Gay Feeling Thermometer t0	58	58	-0.21
Transgender Feeling Thermometer t0	54	53	-0.48
Law Item 1 t0	1.5	1.5	0.78
Law Item 2 t0	1.5	1.4	-1.21
N	912	913	-

Covariate Balance among Pre-Survey Respondents. t-tests.

Table S13.

	Placebo	Treatment	t-statistic
Age	50	49	-0.9
Female	0.54	0.59	1.16
Black	0.27	0.27	-0.14
Gay Feeling Thermometer t0	55	56	0.48
Transgender Feeling Thermometer t0	50	50	0.25
Law Item 1 t0	1.4	1.5	0.63
Law Item 2 t0	1.4	1.3	-0.81
Trans Tolerance Scale t0	-0.01	0	0.11
N	246	255	-

Covariate Balance among Canvassed Respondents.

Table S14.

	Placebo	Treatment	t-statistic
Age	51	49	-1.06
Female	0.57	0.61	0.84
Black	0.25	0.24	-0.24
Gay Feeling Thermometer t0	55	57	0.67
Transgender Feeling Thermometer t0	51	52	0.37
Law Item 1 t0	1.4	1.6	1.31
Law Item 2 t0	1.5	1.4	-0.13
Trans Tolerance Scale t0	0.01	0.04	0.28
N	204	225	-

Covariate Balance among 3-Day Survey Respondents.

Table S15.

	Placebo	Treatment	t-statistic
Age	50	49	-0.62
Female	0.55	0.6	0.88
Black	0.24	0.24	0.04
Gay Feeling Thermometer t0	55	57	0.81
Transgender Feeling Thermometer t0	50	51	0.31
Law Item 1 t0	1.4	1.6	1.01
Law Item 2 t0	1.4	1.4	-0.26
Trans Tolerance Scale t0	0	0.01	0.11
N	194	205	-

Covariate Balance among 3-Week Survey Respondents.

Table S16.

	Placebo	Treatment	t-statistic
Age	50	49	-0.91
Female	0.57	0.59	0.3
Black	0.23	0.23	-0.04
Gay Feeling Thermometer t0	56	57	0.15
Transgender Feeling Thermometer t0	52	50	-0.54
Law Item 1 t0	1.4	1.5	0.52
Law Item 2 t0	1.5	1.4	-0.85
Trans Tolerance Scale t0	0.05	-0.01	-0.53
N	191	210	-

Covariate Balance among 6-Week Survey Respondents.

Table S17.

	Placebo	Treatment	t-statistic
Age	50	50	-0.41
Female	0.58	0.6	0.43
Black	0.25	0.25	0.04
Gay Feeling Thermometer t0	54	57	0.96
Transgender Feeling Thermometer t0	50	52	0.43
Law Item 1 t0	1.4	1.6	1.14
Law Item 2 t0	1.5	1.5	-0.14
Trans Tolerance Scale t0	0.01	0.05	0.39
N	179	206	-

Covariate Balance among 3-Month Survey Respondents.

Table S18.

Condition	Contacted	3 Day	3 Week	6 Week	3 Month
Placebo	255	225	205	210	206
Treatment	246	204	194	191	179

Number of Respondents by Experimental Condition.

Table S19.

Survey Wave	p-Value
3 Day Survey (t1)	0.3233
3 Week Survey (t2)	0.362
6 Week Survey (t3)	0.3376
3 Month Survey (t4)	0.0833

RI p-Value by Survey Wave Test of Differential Attrition by Covariates.

Table S20.

Sample	Female	Age	Republican	Democrat	Af-Am	Hispanic	N
Starting Universe	0.56	49	0.29	0.45	0.3	0.55	68378
Pre Respondent	0.59	47	0.26	0.48	0.25	0.52	1825
Canvassed	0.57	49	0.25	0.49	0.27	0.47	501
3 Day Respondent	0.59	50	0.25	0.48	0.24	0.48	429
3 Week Respondent	0.57	50	0.25	0.47	0.24	0.48	399
6 Week Respondent	0.58	50	0.25	0.47	0.23	0.48	401
3 Month Respondent	0.59	50	0.23	0.5	0.25	0.47	385

Representativeness of Experiment at Each Stage.

Table S21.

	Estimate	Std. Error	t value	p
therm_obama_t1	1.079	2.49	0.433	*n.s.*
therm_obama_t2	-0.991	2.884	-0.344	*n.s.*
therm_obama_t3	-1.155	2.858	-0.404	*n.s.*
therm_obama_t4	1.91	2.743	0.696	*n.s.*
therm_marijuana_t1	3.309	3.005	1.101	*n.s.*
therm_police_t2	0.516	3.101	0.166	*n.s.*
therm_police_t3	0.314	2.926	0.107	*n.s.*
therm_firefighters_t2	3.345	3.124	1.071	*n.s.*
therm_firefighters_t3	1	3.082	0.325	*n.s.*
therm_jbush_t2	-6.846	3.471	-1.973	*n.s.*
therm_jbush_t3	-4.848	3.191	-1.519	*n.s.*
therm_mousavi_t4	0.586	2.538	0.231	*n.s.*
marijuana_pread1	0.016	0.055	0.299	*n.s.*
marijuana_postad	0.022	0.06	0.365	*n.s.*

Reviewer-Suggested Placebo Test. This test shows the treatment had no effect on items that are hypothesized to be unrelated to LGBT-rights, prejudice, or stigma but appeared in the same battery as the transgender feeling thermometer item which showed a large effect.

Table S22.

Year	Mean Feeling Therm
1984	30.9
1988	29.6
1992	38.1
1994	35.2
1996	39.5
1998	45
2000	47
2002	46.1
2004	47.5
2008	49.3
2012	53.5

Average Values of Gay Feeling Thermometer in American National Election Studies, by Year

Table S23.

Variable	Treat Mean	Treat SD	Placebo Mean	Placebo SD
vf_age	50.07	17.62	48.67	17.13
vf_female	0.54	0.5	0.59	0.49
vf_black	0.27	0.45	0.27	0.44
vf_white	0.26	0.44	0.25	0.43
vf_hispanic	0.46	0.5	0.47	0.5
vf_vg_14	0.66	0.47	0.67	0.47
vf_vg_12	0.78	0.42	0.85	0.36
vf_vg_10	0.55	0.5	0.56	0.5
vf_democrat	0.51	0.5	0.46	0.5
vf_republican	0.22	0.42	0.28	0.45
miami_trans_law_t0	1.38	1.74	1.48	1.7
miami_trans_law2_t0	1.42	1.76	1.3	1.67
gender_norm_daugher_t0	0.06	1.19	-0.06	1.18
gender_norm_looks_t0	-0.61	1.26	-0.52	1.22
gender_norm_rights_t0	-0.74	1.18	-0.62	1.16
gender_norms_sexchange_t0	0.29	1.28	0.34	1.14
gender_norms_moral_t0	-0.08	1.32	-0.09	1.24
gender_norms_abnormal_t0	0.1	1.23	0.07	1.17
gender_norm_trans_moral_wrong_t0	-0.04	1.25	-0.04	1.2
ssm_t0	0.33	2.22	0.36	2.14
therm_obama_t0	64.54	35.42	59.78	37.18
therm_gay_t0	54.7	29.63	55.91	27.04
therm_trans_t0	49.71	28.75	50.32	26.45
therm_marijuana_t0	49.75	30.21	46.17	27.82
therm_afams_t0	76.34	23.18	75.31	23.85
ideology_t0	0.02	1.38	-0.01	1.5
religious_t0	2.08	1.76	2.12	1.73
respondent_t0	1	0	1	0
exposure_gay_t0	0.87	0.34	0.89	0.32
exposure_trans_t0	0.3	0.46	0.29	0.45
pid_t0	0.83	2.11	0.55	2.19
scale_for_blocking_t0	-0.11	0.97	-0.1	0.9
miami_trans_law_t1	1.44	1.78	1.57	1.55
miami_trans_law2_t1	1.52	1.7	1.48	1.58
therm_obama_t1	62.04	30.92	57.76	31.51
therm_trans_t1	60.19	29.29	53.72	26.12
therm_marijuana_t1	50.6	26.65	48.36	24.56
gender_norm_looks_t1	-0.37	1.35	-0.54	1.23
gender_norm_rights_t1	-0.7	1.16	-0.65	1.16
gender_norm_sexchange_t1	0.41	1.29	0.35	1.14
gender_norm_moral_t1	0.03	1.34	-0.07	1.17
gender_norm_abnormal_t1	0.15	1.26	-0.05	1.17
gender_norm_trans_moral_wrong_t1	0.28	1.35	0.12	1.17
respondent_t1	0.83	0.38	0.88	0.32
miami_trans_law_t2	1.43	1.79	1.56	1.59
miami_trans_law2_t2	1.46	1.72	1.38	1.62
therm_obama_t2	61.48	30.04	58.04	33
therm_jbush_t2	45.34	26.55	52.56	27.95
therm_trans_t2	58.83	28.15	54.77	26.03
therm_police_t2	63.37	25.36	63.84	24.09
therm_firefighters_t2	79.59	22	77.23	22.11

Variable	Treat Mean	Treat SD	Placebo Mean	Placebo SD
gender_norm_looks_t2	-0.54	1.34	-0.66	1.25
gender_norm_rights_t2	-0.69	1.24	-0.88	1.2
gender_norm_moral_t2	-0.12	1.31	-0.26	1.29
gender_norm_dress_t2	-0.55	1.27	-0.63	1.23
gender_norm_sexchange_t2	0.38	1.28	0.35	1.28
gender_norm_abnormal_t2	0.13	1.26	-0.09	1.24
gender_norm_trans_moral_wrong_t2	0.13	1.26	0	1.26
trans_teacher_t2	0.31	1.26	0.16	1.25
trans_bathroom_t2	0.01	1.22	-0.18	1.24
respondent_t2	0.79	0.41	0.8	0.4
miami_trans_law_withdef_t3	1.39	1.81	1.11	1.78
miami_trans_law2_withdef_t3	1.52	1.74	1.14	1.79
therm_obama_t3	59.35	27.73	56.07	30.34
therm_jbush_t3	45.76	23.46	52.22	24.39
therm_trans_t3	58.4	26.94	52.06	25.65
therm_police_t3	59.22	22.52	59.79	22.08
therm_firefighters_t3	72.16	22.79	71.34	22.04
gender_norm_looks_t3	-0.47	1.24	-0.58	1.28
gender_norm_rights_t3	-0.68	1.12	-0.77	1.16
gender_norm_moral_t3	-0.02	1.27	-0.23	1.31
gender_norm_dress_t3	-0.47	1.25	-0.7	1.24
gender_norm_sexchange_t3	0.55	1.23	0.25	1.26
gender_norm_abnormal_t3	0.21	1.21	-0.09	1.21
gender_norm_trans_moral_wrong_t3	0.23	1.24	0.01	1.23
trans_teacher_t3	0.37	1.24	0.02	1.21
trans_bathroom_t3	0.12	1.26	-0.24	1.23
trans_law_post_ad_t3	0.4	1.81	0.01	1.86
respondent_t3	0.78	0.42	0.82	0.38
nfc_t3	0.08	0.82	0.01	0.83
miami_trans_law_withdef_t4	1.31	1.81	1.12	1.83
miami_trans_law2_withdef_t4	1.4	1.82	1.22	1.76
therm_obama_t4	59.88	25.92	55.56	26.91
therm_trans_t4	58.73	27.65	53.43	26.74
therm_mousavi_t4	42.43	17.27	41.08	18.77
gender_norm_looks_t4	-0.49	1.3	-0.72	1.2
gender_norm_rights_t4	-0.61	1.22	-0.89	1.12
gender_norm_moral_t4	-0.09	1.32	-0.2	1.22
gender_norm_dress_t4	-0.42	1.28	-0.75	1.11
gender_norm_sexchange_t4	0.52	1.18	0.41	1.14
gender_norm_abnormal_t4	0.16	1.25	0.03	1.22
gender_norm_trans_moral_wrong_t4	0.23	1.24	0.03	1.24
trans_teacher_t4	0.36	1.16	0.08	1.2
trans_bathroom_t4	0.08	1.2	-0.31	1.21
respondent_t4	0.73	0.45	0.81	0.39
marijuana_pread1	0.25	0.43	0.22	0.41
marijuana_pread2	1.52	1.58	1.47	1.62
marijuana_postad	0.37	0.48	0.34	0.47
contacted	1	0	1	0
survey_language_es	0.06	0.23	0.07	0.26
sdo_scale	-0.03	0.52	0.01	0.55
treat_ind	1	0	0	0
canvass_trans_ratingstart	7.77	2.78	7.25	3.25

Variable	Treat Mean	Treat SD	Placebo Mean	Placebo SD
canvass_trans_ratingsecond	7.86	2.77	7.62	2.56
canvass_trans_ratingend	8.56	2.33	7.88	2.36
canvass_minutes	9.18	8.49	2.37	3.49
canvasser_trans	0.34	0.48	0.35	0.48
canvasser_id	27.92	16.55	29.68	17.4
cluster_level_t0_scale_mean	-0.08	0.93	-0.09	0.85
vf_independent	0.23	0.42	0.25	0.43
all.dvs.t1	0.12	1.17	0	1
all.dvs.t2	0.1	1.11	0	1
all.dvs.t3	0.25	1.03	0	1
all.dvs.t4	0.15	1.05	0	1
trans.tolerance.dv.t0	-0.01	1.1	0	1
trans.tolerance.dv.t1	0.17	1.14	0	1
trans.tolerance.dv.t2	0.16	1.08	0	1
trans.tolerance.dv.t3	0.29	1.06	0	1
trans.tolerance.dv.t4	0.21	1.05	0	1
miami_trans_law_t0_avg	1.4	1.64	1.39	1.58
miami_trans_law_t1_avg	1.48	1.67	1.52	1.47
miami_trans_law_t2_avg	1.45	1.68	1.47	1.5
miami_trans_law_t3_avg	1.45	1.72	1.13	1.71
miami_trans_law_t4_avg	1.35	1.73	1.17	1.73
gender_nonconformity_t0	-0.1	1.06	0	1
gender_nonconformity_t1	0.05	1.04	0	1
gender_nonconformity_t2	0.12	1.08	0	1
gender_nonconformity_t3	0.14	1.02	0	1
gender_nonconformity_t4	0.28	1.14	0	1
treatment.delivered	0.78	0.41	0.04	0.2
t1.resid	0.11	0.58	-0.1	0.49
transformed.outcome	0.22	1.15	0.2	0.99

Mean and Standard Deviation for All Variables. Note that for the indices placebo group mean is 0 and standard deviation is 1 by construction.

Table S24. Codebook

Variable	Description
id	Respondent ID
vf_age	Voter File Age
vf_party	Voter File Party
vf_racename	Voter File Race
vf_female	Voter File Female
vf_black	Voter File AfAm
vf_white	Voter File White
vf_hispanic	Voter File Hispanic
vf_vg_14	Voter File Voted in 2014
vf_vg_12	Voter File Voted in 2012
vf_vg_10	Voter File Voted in 2010
vf_democrat	Voter File: Registered Democrat
vf_republican	Voter File: Registered Republican
miami_trans_law_t0	See Outcomes section
miami_trans_law2_t0	See Outcomes section
gender_norm_daugher_t0	Parents usually maintain stricter control over their daughters than their sons
gender_norm_looks_t0	See Outcomes section
gender_norm_rights_t0	See Outcomes section
gender_norms_sexchange_t0	See Outcomes section
gender_norms_moral_t0	See Outcomes section
gender_norms_abnormal_t0	See Outcomes section
gender_norm_trans_moral_wrong_t0	See Outcomes section
ssm_t0	Pre-survey, same sex marriage support
therm_obama_t0	Pre-survey, Obama feeling therm
therm_rubio_t0	Pre-survey, Rubio feeling therm
therm_gay_t0	Pre-survey, gay men feeling therm
therm_trans_t0	See Outcomes section
therm_marijuana_t0	Pre-survey, marijuana user feeling therm
therm_afams_t0	Pre-survey, AfAm feeling therm
therm_immigrant_t0	Pre-survey, immigrant feeling therm
therm_muslims_t0	Pre-survey, Muslim feeling therm
ideology_t0	Pre-survey, political ideology
religious_t0	Pre-survey, religiousity
respondent_t0	Responded to pre-survey
exposure_gay_t0	Pre-survey, knows gay people
exposure_trans_t0	Pre-survey, knows trans people
pid_t0	Pre-survey, partisanship
scale_for_blocking_t0	Pre-survey, factor used for blocking
miami_trans_law_t1	See Outcomes section
miami_trans_law2_t1	See Outcomes section
therm_trans_t1	See Outcomes section
gender_norm_looks_t1	See Outcomes section
gender_norm_rights_t1	See Outcomes section
gender_norm_sexchange_t1	See Outcomes section
gender_norm_moral_t1	See Outcomes section
gender_norm_abnormal_t1	See Outcomes section
gender_norm_trans_moral_wrong_t1	See Outcomes section
respondent_t1	3 Day Survey Respondent
miami_trans_law_t2	See Outcomes section
miami_trans_law2_t2	See Outcomes section
therm_trans_t2	See Outcomes section

Variable	Description
gender_norm_looks_t2	See Outcomes section
gender_norm_rights_t2	See Outcomes section
gender_norm_moral_t2	See Outcomes section
gender_norm_dress_t2	See Outcomes section
gender_norm_sexchange_t2	See Outcomes section
gender_norm_abnormal_t2	See Outcomes section
gender_norm_trans_moral_wrong_t2	See Outcomes section
trans_teacher_t2	See Outcomes section
trans_bathroom_t2	See Outcomes section
respondent_t2	3 Week Survey Respondent
miami_trans_law_withdef_t3	See Outcomes section
miami_trans_law2_withdef_t3	See Outcomes section
therm_trans_t3	See Outcomes section
gender_norm_looks_t3	See Outcomes section
gender_norm_rights_t3	See Outcomes section
gender_norm_moral_t3	See Outcomes section
gender_norm_dress_t3	See Outcomes section
gender_norm_sexchange_t3	See Outcomes section
gender_norm_abnormal_t3	See Outcomes section
gender_norm_trans_moral_wrong_t3	See Outcomes section
trans_teacher_t3	See Outcomes section
trans_bathroom_t3	See Outcomes section
trans_ad_displayed_t3	Did the opposing message ad properly play?
trans_law_post_ad_t3	See Outcomes section
respondent_t3	6 Week Survey Respondent
nfc_t3	Need for Cognition Scale
marijuana_pread1	Support legalizing marijuana
marijuana_postad	Support legalizing marijuana
miami_trans_law_withdef_t4	See Outcomes section
miami_trans_law2_withdef_t4	See Outcomes section
therm_trans_t4	See Outcomes section
gender_norm_looks_t4	See Outcomes section
gender_norm_rights_t4	See Outcomes section
gender_norm_moral_t4	See Outcomes section
gender_norm_dress_t4	See Outcomes section
gender_norm_sexchange_t4	See Outcomes section
gender_norm_abnormal_t4	See Outcomes section
gender_norm_trans_moral_wrong_t4	See Outcomes section
trans_teacher_t4	See Outcomes section
trans_bathroom_t4	See Outcomes section
respondent_t4	3 Month Survey Respondent
exp_actual_convo	Actual Treatment Delivered (as opposed to assigned)
contacted	Voter came to door
survey_language_es	Survey conducted in Spanish
sdo_scale	Pre-survey, social dominance orientation scale
treat_ind	Treatment assignment
canvass_trans_ratingstart	In canvass, rating of anti-discrimination law
canvass_minutes	Length of canvass conversation
canvasser_experience	Canvasser previous experience
canvasser_trans	Canvassed identified as transgender or gender non-conforming
canvasser_id	Canvasser identifier
hh_id	Household identifier

Variable	Description
block_ind	Block Identifier
cluster_level_t0_scale_mean	Scale used for blocking at household level
survey_language_t0	Survey language

Table S25. AAPOR standard definition response rates.

RR	t0 (Baseline)	t1	t2	t3
RR1	2.67%	84.6%	79.6%	79.0%
RR2	2.82%	85.6%	81.6%	80.0%
RR3	2.67%	84.7%	79.7%	79.1%
RR4	2.82%	85.7%	81.7%	80.1%