From Extreme to Mainstream:
The Erosion of Social Norms[*]

Leonardo Bursztyn[†]
Georgy Egorov[‡]
Stefano Fiorin[§]

March 2020

Abstract

Social norms, usually persistent, can change quickly when new public information arrives, such as a surprising election outcome. People may become more inclined to express views or take actions previously perceived as stigmatized and may judge others less negatively for doing so. We examine this possibility using two experiments. We first show via revealed preference experiments that Donald Trump's rise in popularity and eventual victory increased individuals' willingness to publicly express xenophobic views. We then show that individuals are judged less negatively if they publicly expressed a xenophobic view in an environment where that view is more popular.

Keywords: Social norms; social acceptability; elections; xenophobia; political attitudes; social interactions; communication

1 Introduction

Social norms, the set of 'social sanctions or rewards' that incentivize a certain behavior (Bénabou and Tirole, 2011), are an important element of any society: some behaviors and opinions are socially desirable, while others are stigmatized. There is growing evidence that individuals care to a large extent about how they are perceived by others and that such concerns might affect important decisions in a variety of settings, from charitable donations (Andreoni and Bernheim, 2009; DellaVigna, List and Malmendier, 2012; Andreoni, Rao and Trachtman, 2017) to schooling choices (Bursztyn and Jensen, 2015) to political behavior (Gerber, Green and Larimer, 2008; DellaVigna et al., 2017; Enikolopov et al., 2017; Perez-Truglia and Cruces, 2017). Moreover, these social image concerns matter both in interactions with other people from the same social group (Bursztyn and Jensen, 2015) and in interactions with strangers, such as surveyors and solicitors (DellaVigna, List and Malmendier, 2012; DellaVigna et al., 2017).

A recent literature has documented the persistence of cultural traits and norms over long periods of time (Voigtländer and Voth, 2012; Fernández, 2007; Giuliano, 2007; Algan and Cahuc, 2010; Alesina, Giuliano and Nunn, 2013). However, little is known about what factors might lead long-standing social norms to *change*, or even more so, to change *quickly*. In this paper, we argue that aggregators of private opinions in a society, such as elections, might lead to updates in individuals' perceptions of what people around them think, and thus induce fast changes in the social acceptability of holding and expressing certain opinions. This, in turn, might increase the likelihood that these opinions are publicly expressed and reduce the extent to which these expressions are negatively judged and sanctioned by others.[1]

Consider the support for the communist regime in the Soviet Union in the late 1980s. Kuran (1991) argues that many individuals opposed the regime but believed that others supported it. In that environment, a referendum on the regime would have quickly updated people's opinions about the views of others. Incorrect beliefs about the opinions of others are not restricted to totalitarian regimes, where expressing personal views is often risky. In fact, as we argue below, if most individuals assume that a specific opinion is stigmatized, the stigma may be sustained in equilibrium.[2]

---

[1]In the view of social norms we adopt, they guide public or potentially public, but not private actions. We do not take a broader view on social norms that also includes self-image concerns that can shape even one's private behavior by rewarding adherence and punishing deviance. Notice that since our paper explains how new public information can change social norms in the narrower (and our preferred) sense, it also suggests that social norms in the broader sense change – specifically, the part of social norms responsible for rewards or punishments by *others*. Thus, when we say that social norms are eroded, they do so according to either definition.

[2]This phenomenon is known in social psychology as "pluralistic ignorance" (Katz and Allport, 1931), where privately most people reject a view, but incorrectly believe that most other people accept it, and therefore end up acting accordingly. For example, in 1968 most white Americans substantially overestimated the support for racial segregation among other whites (O'Gorman, 1975). A related concept is "preference falsification" (Kuran, 1995): people's stated, public preferences are influenced by social acceptability, and might be different from their true, private preferences. For example, American college graduates consistently understate their support for immigration

In this paper, we examine how social norms can be eroded quickly when new public information arrives naturally, such as an election outcome.[3] We use experiments to test the idea that Donald Trump's rise in popularity and eventual victory in the 2016 U.S. Presidential election causally increased individual's willingness to publicly express anti-immigrant (xenophobic) views, as well as the social acceptability of such expression.[4] We examine both sides of a social interaction: agents engaging in observable behavior and observers, judging and potentially sanctioning/rewarding the behavior of agents. Our first experiment indicates that Donald Trump's rise in popularity and eventual victory in the 2016 U.S. changed increased individuals' willingness to engage in public xenophobic behavior, a behavior previously believed to be stigmatized. We also examine how this process led to changes on the *observer*'s side as well. We document that even individuals likely disagreeing with xenophobic behavior judge less negatively (and therefore sanction less) public expression of xenophobia when they learn that the underlying view is more popular by observing the election outcome. This paper therefore studies a *natural* process of social norm *change*, and it does so examining both agents and observers.

To organize thoughts and connect the experiments within a single framework, we build a model where two types of individuals, say xenophobic and tolerant, choose an action, but in doing so they care about approval or disapproval of other people who might observe the action. Like the agents choosing the action, the members of the audience are Bayesian, and their inference about the agent's type depends on the strategies he uses in equilibrium. In this environment, social pressure might lead some agents to choose the action that they do not naturally prefer, and it is even possible that all agents choose the same action (which, arguably, prevents learning about the distribution of types).[5] We then study the impact of public signals and show that a signal suggesting that more people are likely to be xenophobic increases the share of agents who choose the xenophobic action. However, the same signal may decrease the audience's perception that an agent who chose the xenophobic action is a xenophobe. Indeed, when few individuals are perceived

to be xenophobic, there is no social pressure to appear to be one, and thus only xenophobes would choose a public xenophobic action. In contrast, when xenophobic individuals are thought to be common, such social pressure might be there, and thus not everyone acting in a xenophobic way is a true xenophobe.[6]

We experimentally capture the effects of Trump's rise in popularity. Throughout his campaign, Donald Trump proposed, among other things, the construction of a wall separating the U.S. and Mexico and a ban on Muslims from entering the U.S. His popularity might thus send an informative signal about the number of people who sympathize with these proposals and thus about those who hold xenophobic views. As a result, Donald Trump's electoral success potentially caused a shift in social norms regarding expressing views on immigrants. We first examine the effect of Trump's rise in popularity (and thus of information aggregation) on people's willingness to publicly express xenophobia. In August-October 2018, we recruited a sample of 1,600 participants through an online panel survey company. We manipulate respondents' perceptions of Trump's local level of popularity in the 2016 election by exploiting the fact in some areas of the U.S. where that election was close, the candidate who won the election at the county level was different from the winner at the metropolitan statistical area (MSA) level. The subjects of this experiment were all recruited from the Pittsburgh MSA. In the beginning of the experiment, participants were given three facts about the history and politics of this area; we randomize whether one of these facts was that Donald Trump won the election in "Pittsburgh's metropolitan area" or that Hillary Clinton won the election in "Pittsburgh's county." Participants were then offered a bonus cash reward if they authorized the researchers to make a donation to a strongly anti-immigration organization on their behalf. Accepting the offer is therefore a *profitable* xenophobic action. We also randomly manipulated the participants' expectations of anonymity through a public condition that implies visibility by a relevant "peer group": other subjects in the same geographic area of the respondent. All participants were told that the donation decisions would be posted on a website to be shared with all the participants from their area. Half of the participants were assured that their individual responses would be kept completely anonymous (the "private" condition). The other half, instead, were exposed to a treatment designed to make them think that the responses posted on the website would not be anonymous (the "public" condition).

In the "Clinton won" treatment, participants in the public condition were significantly more likely to forgo the donation bonus payment than those in the private condition. This suggests the presence of social stigma associated with the action. However, we find that for the participants in the "Trump won" treatment, who received information that positively updated their perceptions of Trump's popularity, the wedge in the likelihood of undertaking the xenophobic action in private

---

[6]A similar effect appears in Bénabou and Tirole (2006), where an explicit incentive to do a prosocial action enables selfish people to do so, thus reducing the signaling value of such action. Similarly, in Ali and Lin (2013), a higher propensity of opportunistic voters (those caring about social image but not social welfare) to vote makes other such voters less willing to turn out.

and public disappeared. This difference with respect to the "Clinton won" condition was driven entirely by an increase in the donation rate in the public condition, with no change in the private condition. Our results suggest that an increase in perceptions of Donald Trump's local popularity does not make these participants more xenophobic, but instead makes those who were already xenophobic more comfortable expressing their xenophobic views in public. In addition, consistent with an underlying mechanism of updates in perceptions about the local popularity in xenophobic views, we find that the "Trump won" treatment shifts the distribution of participants' perceptions about the local popularity of those views to the right. We also use the "Trump won" treatment as an instrument for the shift in perceptions, and show that participants' perceptions causally affect their donation in public, but not in private. Note that instead of directly manipulating people's beliefs about a social norm (e.g., by telling them the share of people who hold a certain view), we instead vary the perceived local popularity of Donald Trump, and document that this naturally leads to changes in behavior.

Overall, these results indicate that social norms can quickly shift as a result of private opinions being aggregated and diffused through elections. In an earlier, motivating experiment implemented on *Amazon Mechanical Turk (mTurk)*, we took advantage of a unique opportunity and examined information aggregation in "real time" in the weeks just before and after the November 2016 U.S. Presidential election. The results are consistent with the causal effect we document in the main experiment. Participants who expected their decision to donate to a xenophobic organization to be observed by a surveyor were significantly less likely to make the donation than those expecting anonymity. Increases in participants' perceptions of Trump's popularity (either through experimental variation or through the "natural experiment" of his victory) eliminated the wedge between private and public behavior. The design and results from the motivating experiment are presented in Appendix D.

We then turn our attention to the study of how the 2016 election changed people's interpretation of observed xenophobic expression and their decisions to sanction it. We again hired an online panel survey company to recruit, in November 2018, a sample of about 1,800 respondents who had previously self-identified to the firm as Democrats. The respondents are asked to play a dictator game in which they decide how to split $2 between themselves and another agent. Our goal is to evaluate how negatively they judge (and how they punish) someone for expressing xenophobia in private or in public, depending on the perceptions of the popularity of xenophobia in that person's area. We tell a recruited participant (henceforth *player 1*) that a participant from a previous study (*player 2*) donated money to a named xenophobic organization, and cross-randomize whether we tell player 1 that: i) Trump or Clinton won the election in player 2's area; and ii) player 2's donation was in private or publicly observable on a website shared with other participants from his area. Note that participants assigned as player 2 were real participants from the experiment described above. Other design elements are important to emphasize. Player 2 did not know he was going to

be part of a dictator game in the future and player 1 was informed about this fact. Also, player 1 did not know that player 2's area was Pittsburgh (doing otherwise would have potentially limited the size of the shifts in player 1's perceptions about Trump's popularity in player 2's area). Finally, player 1 was not told that player 2 had been offered an incentive to donate to the organization.

We find that the amount given by player 1 participants to player 2 participants are similar when player 2's donation was in private, regardless of who won the election in player 2's area. This helps us deal with confounds, such as learning about local motives to privately oppose immigration (since shifting perceptions on who won the election in player 2's area could have changed the perception of what that area is). Player 1 participants also give a similar amount to player 2 when his decision was in public when told that Clinton won in player 2's area. The logic is simple: there is no social pressure to act as a xenophobe in an area where Clinton won (the pressure is, if anything, in the opposite direction). Therefore it must be that player 2 is indeed a xenophobe, just like in the private cases that also did not involve social considerations. In the final treatment, where player 2's decisions were to be observed in an area where Trump won the election, player 1 participants give significantly more to player 2. Player 2 was now potentially subject to social pressure to publicly act as a xenophobe since his donation decision was going to be observed by local peers in an area where Trump won the election. The inference player 1 could make on whether player 2 was truly a xenophobe is therefore weakened due to the strategic incentives player 2 could have to donate in order to pander to the majority. Here again, we use the information on who won the election in player 2's area as an instrument for player 1's perception of the popularity of xenophobia in that area, and show that these perceptions affect the amount given by player 1 when player 2's decision was public but not when it was private.

In Appendix E we present the design and results from a similar experiment conducted on *mTurk*, where participants were asked to play a dictator game with another respondent in Switzerland. In that experiment, we manipulated perceptions of the popularity of anti-Muslim sentiment in Switzerland, by randomly giving information about the 2009 Swiss referendum that banned the construction of minarets in that country. The results are again consistent with the predictions of our framework.

Our results suggest that Trump's rise in popularity and eventual electoral victory casually changed social norms regarding the expression of xenophobic views in the U.S. Though we detect no changes in *privately-held* views, we believe the findings on public expression are of great policy relevance.[7]

---

[7]For example, increases in public expression of anti-immigrant sentiment might also lead to more frequent acts of hate crime against immigrants, and might also facilitate coordination for large-scale actions, such as demonstrations and movements. Recent work provides evidence that such demonstrations and movements might affect many important outcomes, from election results (Madestam et al., 2013) to the stock market valuation of different firms (Acemoglu, Hassan and Tahoun, 2018). In addition, a reduction in the stigma associated with holding previously-extreme views might lead to shifts in the language used in and reported by the popular media, and might also reduce the stigma associated with consuming and discussing certain news sources on the far side of the political spectrum.

Related Literature. Our results contribute to a growing literature that examines the impacts of political institutions on social norms and culture more generally. This literature typically studies the long-run impact of political institutions (e.g., Lowes et al., 2017); we show that changes on the political side can lead to fast changes in social norms. Our paper also adds to a recent theoretical literature on social norms (e.g., Bénabou and Tirole, 2011; Acemoglu and Jackson, 2017) by studying how new information may lead to unraveling of such norms. Our findings also speak to a cross-disciplinary literature on the consequences of political actions, both theoretical (e.g., Lohmann, 1993) and empirical (e.g., Madestam et al., 2013).[8]

Our work also relates to existing papers studying the economic consequences of conformity. Prendergast (1993) identifies rational incentives for managers to conform to supervisors' opinions in order to appear competent, which in turn hampers information transmission, while Bernheim (1994) shows that social concerns can lead to formation of social norms. Developing these ideas, Morris (2001) shows that the fear of being seen as biased could completely shut down information transmission from an advisor to a decision-maker. Ali and Bénabou (2016) and Ali and Lin (2013) study how social image concerns can give rise to pro-social behavior by non-altruistic individuals, in the contexts of contributions to public goods and voter turnout, respectively.[9] Andreoni, Nikiforakis and Siegenthaler (2017) study 'conformity traps,' situations where groups of individuals fail to coordinate on a beneficial action due to individual incentives to conform to the predominant and inefficient behavior. In a laboratory experiment they find, in particular, that opinion polls can facilitate changes of norms that benefit the group. Their setting, however, is one of full information, and thus opinion polls facilitate switching from one equilibrium to another. Our model has incomplete information and features a unique equilibrium, and elections can change the beliefs about the distribution of other people's opinions (though we do not take a position on whether overcoming conformity is necessarily socially beneficial).

Our paper also contributes to a recent experimental literature on the effect of social norms on behavior. Krupka and Weber (2013) show that elicited social norms predict changes in behavior across variants of the dictator game. Bursztyn, González and Yanagizawa-Drott (2018) directly manipulate perceived social norms, which in turn changes behavior. Here, our interest is in evaluating how natural processes aggregating information about private opinions (such as an election)

An increase in public expression of such views can thus lead to an increase in individuals' overall exposure to them, and more exposure might eventually lead to changes in privately-held views, via persuasion or simple conformism.

[8]Recent work has documented how policy changes, such as the introduction of cable TV in India (Jensen and Oster, 2009), and quotas for female politicians (Beaman et al., 2009) can lead to changes in attitudes toward women and in behavior. However, these studies do not focus on isolating the mechanisms of social norm change (i.e., updates in perceived social acceptability as opposed to potential changes in private views.)

[9]Our model builds on these important precursors, albeit with notable differences. In these models, like in our paper, increased propensity of non-altruistic individuals to act prosocially makes it more difficult for observers to identify truly prosocial individuals. The main difference is that we seek to capture a contentious political setting where the same social image (e.g., of a xenophobic person) might be approved by some and stigmatized by others, as opposed to situations where all individuals seek to espouse a particular pro-social image.

can lead to fast changes in social norms which lead both to changes in behavior *and* in sanctions of behavior. Our approach also allows us to evaluate, using revealed preference, how updates about existing norms change both the way people express themselves and the way people judge and sanction others for expressing themselves.

The remainder of this paper proceeds as follows. We introduce a simple framework formalizing our argument in Section 2. In Section 3, we present the design and results from Experiment 1, which studies the expression of xenophobic views. In Section 4, we present the design and results from Experiment 2, which studies the judgment and sanctioning of xenophobic expression. Section 5 concludes.

2 Motivating Framework

To organize thoughts and motivate our experimental designs, we present a simple model of communication.

2.1 Model

A society S consists of individuals that may have one of two types, A or B (so $t_i \in \{A, B\}$ for citizen i). The distribution of types is i.i.d, with $\Pr\{t_i = A\} = p$. This probability p is itself a random variable with realization chosen by Nature at the beginning of the game from two possible values, so that $p = p_H$ with probability θ and $p = p_L$ with probability $1 - \theta$, where $\theta \in (0, 1)$ and $0 < p_L < p_H < 1$. Each citizen knows his/her own type but not the realization of p (but might get signals about that); however, the process by which p is determined is known. In what follows, we will refer to the citizen choosing an action (the actor) as "he" and to a generic member of the audience as "she."

The actor needs to decide between two actions, which we also denote A and B, slightly abusing notation. We interpret action A as the preferred action of type A and B as the preferred action of type B. More specifically, assume that the utility levels of each type from the corresponding action are $V_A > 0$ and $V_B > 0$, respectively, and the utility levels from the opposite actions are normalized to zero.

Suppose that citizen i is the actor choosing $d_i \in \{A, B\}$, and suppose he is doing so before an audience of size $n \geq 0$, with $n_{priv} = 0$ corresponding to a private decision and $n_{pub} > 0$ to a public decision (in which case let N_i denote this audience). Members of the audience observe the decision d_i and use it, as well as any other information they have, to update their beliefs on citizen i's type t_i. We assume that citizen i gets utility χ_i if type A member of the audience believes he is type A (as opposed to B) or if type B one believes he is type B (as opposed to A).[10] Thus, citizen i

---

[10]This reputational payoff is exogenously given in the model, but one can easily see how it may be endogenized, and how it may even get an instrumental component. In the Appendix A.5, we model the behavior of a receiver who

makes decision d_i to maximize

$$
\begin{aligned}
U_i\left(d_i\right) & = V_{t_i}\mathbf{I}\left\{d_i = t_i\right\} + \mathbf{I}_{pub}\chi_i \sum_{j \in N_i} \mathbf{E}_i\left(\Pr_j\left(t_i = t_j \mid d_i, t_j\right) \mid t_i\right) \\
& = V_{t_i}\mathbf{I}\left\{d_i = t_i\right\} + \mathbf{I}_{pub}h_i \sum_{t \in \{A,B\}} \Pr_i\left(t_j = t \mid t_i\right)\Pr_j\left(t_i = t \mid d_i, t_j = t\right),
\end{aligned}
$$

where $h_i = n_{pub}\chi_i$ denotes individual i's total social image concerns when making a public decision and j is a generic member of the audience. We will use this notation in what follows.

We are interested in Perfect Bayesian equilibria that satisfy the D1 criterion (Cho and Kreps, 1987). Throughout the paper we assume that social image concerns h_i are distributed smoothly and that direct payoffs from preferred action (V_A and V_B) are not too small relative to these social image concerns.

Assumption 1. *Social image concerns h_i are independent of type t_i and are distributed on $[m, M]$ with c.d.f. $F(h)$ and p.d.f. $f(h)$ such that $f(h)h\left(h - \tilde{V}\right) \leq \tilde{V}$, where $\tilde{V} = \min\{V_A, V_B\}$. In addition, $\tilde{V} \geq (p_H - p_L)M$.*[11]

can share with the sender in a dictator game, and argue that modeling receivers as having homophily in addition to altruism predicts that the sender's payoff is monotonically increasing in the receiver's posterior about the sender's type. Out modeling choice maps well into our second experiment, and is consistent with existent empirical evidence. For instance, Hoffman, McCabe and Smith (1996) document that social distance decreases transfers in dictator games, and Fowler and Kam (2007) show that donations to individuals of a different political orientation in a dictator game are lower. More broadly, perception of proximity of preferences between the two parties could have other material consequences, for example, by making credible communication possible in a Crawford and Sobel (1982) style model.

[11]The first part of the assumption rules out multiplicity of monotonic equilibria, i.e., those where senders of type A are more likely to choose action A than senders of type B. To see why such multiplicity is possible, suppose that more senders of type A start choosing action B. Then receivers of both types increase their posterior belief that conditional on choosing B, the type is A. However, because these posteriors are different (the receivers take their own type realization into account), it is well possible that receivers of type A are more sensitive to this change, for example because type B receivers have a very high prior that most senders are their type. Now, if type B receivers are barely affected by more A type senders choosing B, whereas type A receivers now have a higher posterior that the sender who chose action B might still be type A, this creates incentives to more type A senders to choose action B. The assumption that there is sufficient heterogeneity in social image concerns is made to ensure that this strategic complemenarity effect is not strong enough. Notice that such multiplicity may arise because *receivers* of different types have different beliefs about the distribution of *senders*.

The second part of the assumption rules out counterintuitive nonmonotone equilibria where individuals of type A choose action B and individuals of type B choose action A, and neither type wants to deviate because they believe (due to the realization of their own type) that their type is more numerous. They therefore want receivers to make the correct inference, which creates the incentive to avoid mimicking the opposite type, which in such equilibrium requires sticking to the opposite action. The assumption that the direct benefit from the action is not too small rules this out. Notice that unlike the previous case, this type of multiplicity may arise because *senders* of different types have different beliefs about the distribution of *receivers*.

In the absence of assumption 1 there may be multiple equilibria, as Example 1 and 2 in the Appendix A.2 demonstrate. However, for monotonic equilibria (arguably the more intuitive), the comparative statics results hold more generally, in the Milgrom and Shannon (1994) sense (e.g., for minimal or maximal equilibrium).

2.2 Analysis

Each individual i, upon learning his/her type, updates on the distribution of types in the society. For type A and type B the expected shares of type A equal, respectively,

$$
\begin{aligned}
P_A &= \frac{\theta p_H^2 + (1-\theta)\, p_L^2}{\theta p_H + (1-\theta)\, p_L}; \\
P_B &= \frac{\theta p_H\,(1-p_H) + (1-\theta)\, p_L\,(1-p_L)}{\theta\,(1-p_H) + (1-\theta)\,(1-p_L)}.
\end{aligned}
$$

Proposition 1. *Under Assumption 1 there is a unique equilibrium. In the private decision case, citizen i chooses the action that corresponds to his type, $d_i = t_i$. In the public decision case:*

(i) If $V_B \leq mP_B\,(P_A + P_B - 1)$, then citizen i chooses $d_i = A$;

(ii) If $V_B \in (mP_B\,(P_A + P_B - 1), M\,(2P_B - 1))$, then citizen i of type A chooses $d_i = A$, while citizen of type B chooses $d_i = A$ if h_i is above some threshold and $d_i = B$ otherwise;

(iii) If $V_B \geq M\,(2P_B - 1)$ and $V_A \geq M\,(1 - 2P_A)$, then citizen i chooses $d_i = t_i$;

(iv) If $V_A \in (m\,(1 - P_A)\,(1 - P_A - P_B), M\,(1 - 2P_A))$, then citizen i of type B chooses $d_i = B$, while citizen of type A chooses $d_i = B$ if h_i is above some threshold and $d_i = A$ otherwise;

(v) If $V_A \leq m\,(1 - P_A)\,(1 - P_A - P_B)$, then citizen i chooses $d_i = B$.

These cases are exhaustive and mutually exclusive. Moreover, an increase in θ weakly increases the share of individuals of either type who choose $d_i = A$, and strictly so for type B in (ii) and for type A in (iv).

Let us introduce an informative public signal s about the share of type A in the society (which was denoted by p) that can take two values: $s \in \{p_L, p_H\}$. Assume that $s = p$ with probability $\mu \geq \frac{1}{2}$ ($\mu = \frac{1}{2}$ corresponds to an uninformative signal, and $\mu = 1$ to a precise revelation of p). Signal s changes the perception about the distribution of types in the society from the standpoint of any agent. For an outsider (or someone who has not learned their type), the probability that $p = p_H$ is given by

$$
\theta(s) = \begin{cases} \frac{\theta\mu}{\theta\mu + (1-\theta)(1-\mu)} & \text{if } s = p_H \\[2mm] \frac{\theta(1-\mu)}{\theta(1-\mu) + (1-\theta)\mu} & \text{if } s = p_L \end{cases}.
$$

For any member of the society, posteriors will depend on their type, and the formulas for P_A and P_B remain valid if we replace θ with $\theta(s)$.

This signal both changes citizen i's belief about the distribution of types in the audience and also changes the way the audience updates about citizen i's type. Thus, citizen i is now maximizing

$$
U_i\,(d_i; s) = V_{t_i}\mathbf{I}\{d_i = t_i\} + \mathbf{I}_{pub} h_i \sum_{t \in \{A,B\}} \Pr_i\,(t_j = t \mid t_i, s)\Pr_j\,(t_i = t \mid d_i, t_j = t, s).
$$

Here, the first instance of s corresponds to audience members' posterior that the citizen is type A

(B) being higher if the signal is $s = p_H$ ($s = p_L$); the second instance captures the citizen believing that audience is more likely to consist of A types (B types) if $s = p_H$ ($s = p_L$). Both effects push the citizen towards choosing action A if $s = p_H$ (the intuition for $s = p_L$ is similar). The latter effect is straightforward: if the audience is more likely to consist of A types, then choosing action A is more likely to boost one's social image from their viewpoint. The former effect is slightly subtler: if a member of the audience has a strong prior that the citizen is type A, it is difficult for a citizen of type A to persuade her otherwise, and he might as well give up on the social image concerns and choose the action he likes.

The above intuition is summarized in the following proposition.

Proposition 2. *A citizen is (weakly) more likely to choose $d_i = A$ if he receives signal $s = p_H$ as compared to the case of no signal. If he receives signal $s = p_L$, the citizen is (weakly) less likely to choose $d_i = A$ than in the case of no signal.*

Let us now analyze how signal s affects the posterior probability that the citizen who chose action A is indeed type A. Let

$$P_{B,s} = \frac{\theta\left(s\right) p_H \left(1 - p_H\right) + \left(1 - \theta\left(s\right)\right) p_L \left(1 - p_L\right)}{\theta\left(s\right)\left(1 - p_H\right) + \left(1 - \theta\left(s\right)\right)\left(1 - p_L\right)}$$

denote the posterior of type B individual about the share of type A, conditional on receiving signal s.

Proposition 3. *Suppose that $V_B > M\left(2P_{B,p_L} - 1\right)$, so for low signal $s = p_L$, citizens of type $t_i = B$ choose action $d_i = B$. Then the audience's posterior that a citizen who chose action A is indeed type A is (weakly) lower for high signal $s = p_H$ than for low signal $s = p_L$. (The converse is true for the posterior probability that the citizen has type B.)*

In other words, if for signal $s = p_L$ indicating that type B is relatively frequent, citizens of type B choose the corresponding action, then anyone who chose action A must be indeed type A, i.e. the posterior is equal to 1. A signal $s = p_H$ indicating that type A is more frequent might lead to some B type citizens picking action A, making it possible that the citizen who chose action A is actually type B, making the posterior lower than 1. Of course, in the absence of an audience, i.e., in the private case, this posterior equals 1 for either signal, which is the same as in the public case with signal $s = p_L$; in contrast, the public case with $s = p_H$ is the only one in which this posterior may be lower than 1.

In Appendix A.3, we provide a numerical example and a figure to illustrate our theoretical predictions.

3 Experiment 1: Expressing Xenophobia

3.1 Experimental Design

Between August and October 2018, an online panel survey company recruited participants (N=1,600) from the seven counties composing the Pittsburgh, PA Metropolitan Statistical Area (Allegheny, Armstrong, Beaver, Butler, Fayette, Washington, Westmoreland).[12] Each panelist could participate in the survey only once.[13]

After participants answered a number of demographic questions, we randomized the perceptions of Trump's popularity among the respondents' peers using an innovative strategy that exploited the variation in the identity of the winner of the popular vote in the 2016 presidential election across overlapping geographical areas. While Donald Trump won the popular vote in Pittsburgh's metropolitan area, Hillary Clinton won it in Allegheny, the county where Pittsburgh is located. At the beginning of the survey, respondents were told three facts about Pittsburgh politics. Half of the participants were randomized into the *Trump won* condition, and had one of the facts state that:

> "In the 2016 US Presidential Election, Donald Trump won Pittsburgh's metropolitan area."

The other half of the participants were instead randomized into the *Clinton won* condition, and were told that:

> "In the 2016 US Presidential Election, Hillary Clinton won Pittsburgh's county."

The remaining two bullet points were the same for both groups of participants and contained neutral historical information about Pittsburgh politics. These two extra bullets were included to limit the participants' ability to infer the purpose of the study.

In the next part of the intervention, we measured the perceived social acceptability of strong anti-immigrant sentiment using a donation experiment with real stakes. Participants were first told that they would be given the opportunity to make a donation to a randomly drawn organization that could either be anti- or pro-immigration. Before the organization was revealed, the participants were also told that the donation would not be subtracted from their payment for participating in the survey, and that in case they authorized the donation they would be paid an extra $1. We

---

[12]The company, Qualtrics, is the same used by Elias, Lacetera and Macis (2016) and works as a panel aggregator that leverages a variety of actively managed participant databases. The databases consist of online participants who have opted into participating in market research studies in exchange for incentives. On a quarterly basis, quality evaluations are performed by third-party data specialists to evaluate the databases on key aspects including response quality and consistency of the panelists.

[13]See the survey script in Appendix C.2. The reader can access the survey information sheet and interactively follow the survey at `http://ssd.az1.qualtrics.com/jfe/form/SV_erhJZh0coOf0u6F`.

analyze the behavior of the participants who were asked to the organization we were interested in: the *Federation of American Immigration Reform*. To make sure that the participants were aware of the organization's very strong anti-immigration stance, a few more details about the organization and its founder were provided in the experiment:

> The Federation for American Immigration Reform (FAIR) is an **immigration-reduction organization** of concerned individuals who believe that immigration laws must be reformed, and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of 'The Immigration Invasion' who wrote "I've come to the point of view that for European-American society and culture to persist requires a European-American majority, and a clear one at that."

In addition to the first randomization informing subjects about either Trump's or Clinton's victory, we introduced a second layer of cross-randomization at the donation stage. All of the participants were told that the results of the survey would be posted online. The link to the website was given to all participants, so that it was clear to each participants that other participants (who came from the same geographical area around Pittsburgh, a fact known to the participants) would have access to the result of the survey. Importantly, however, half of the participants were assured that their individual responses would be kept completely anonymous, and that the results would be posted in aggregate form as percentages only: we refer to this condition as the *private* condition. Specifically, participants were told:

> "The anonymized results from this survey will be posted on our website in approximately one month. Results will be reported as percentages only, so your individual response to the survey will remain anonymous. We will notify you when the results become available on our website, [website link]."

The other half of the subjects were instead not given this assurance of anonymity. In practice, their decision appeared on the website in anonymous form as for the other half of the participants. Importantly, however, their decisions were not reported as percentages in aggregated form: instead, for each participant, the website reported the (anonymous) survey response ID and the individual donation decision. To avoid deception, the subjects were never explicitly told that their personal details would be published on the website along with their donation decision. The participants, however, were given clues suggesting that their name and email could be published on the website together with their individual donation decision. This is what we refer to as the *public* condition:

> "The results from this survey, including your individual donation decision and the donation decisions of all of the other Pittsburgh respondents to this survey, will be posted on our website in approximately one month. There is no need to provide your name,

email, etc. here; the survey company we work with has this information already. We
will notify you when the results become available on our website, [website link]."

Note that the survey company indeed had access to the participants' personal identifying infor-
mation, but we (the experimenters) did not. As a result, the statements in both conditions were
factually true.

After the donation decision, participants were asked to predict the share of Pittsburgh voters
that agree with the following anti-immigration statements:

"For European American society and culture to persist requires a European-American
majority, and a clear one at that."

and

"Both legal and illegal immigration should be drastically reduced because immigrants
undermine American culture and do not respect American values."

This provides two measures of the perceived local popularity of anti-immigrant sentiment. At
the end of the survey, the respondents answered a few demographic questions.

Link to Theory.   This experiment looks at the senders' decision. We interpret type A as xeno-
phobic and type B as tolerant; action A as the xenophobic action (authorizing the donation to the
anti-immigration organization) and action B as the tolerant one (not authorizing this donation).
Absent social image concerns, xenophobic individuals should strictly prefer action A (they help an
aligned organization and get a dollar) and tolerant ones should strictly prefer action B (implicitly,
we assume that associating with the organization creates more than a dollar of disutility for tol-
erant people). The citizens are our subjects (survey participants), and the audience (in the public
setting) are those who would visit the website we provided. We interpret information that Trump
won Pittsburgh MSA as the high signal about the share of type A, $s = p_H$, and information that
Clinton won Pittsburgh county as the low signal $s = p_L$.

The model predicts (Proposition 2) that citizens who got signal $s = p_H$ are more likely to choose
the xenophobic action than those who got signal $s = p_L$ in the public setting, and that there is no
difference in the private setting.[14]

---

[14]One caveat is that the model assumes that the same signal is obtained by both the senders and receivers,
which corresponds to survey participants expecting the audience to consist of other survey participants or those
who knew which politician won the election in Pittsburgh area. This is reasonable, as we provided the link to
the survey participants only. However, if they expected the results of the experiment to be available to a larger
Internet audience, including those who did not know who won in the Pittsburgh metropolitan area or county but
who nevertheless mattered for the purposes of social image concerns, this would require a model where senders get
the signal but receivers do not. As we show in Appendix A.4, the results of Proposition 2 continue to hold as stated.
Thus, the predictions of our framework are robust to our subjects' interpretation of who exactly will get access to
the results of the survey.

3.2 Main Results

Figure 1: **Experiment 1: Donation Rates**
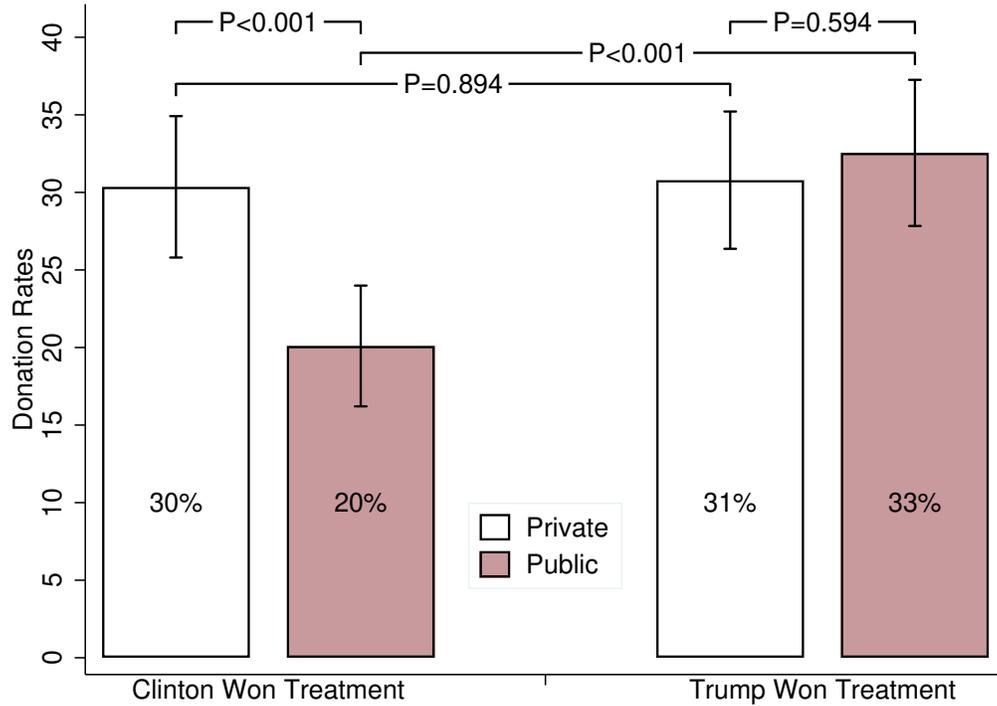
Notes: the two bars on the left display donation rates to the anti-immigration organization for individuals in the private and public conditions in the Clinton Won treatment group (respectively N=392 and N=408), and the two bars on the right display those for individuals in the private and public conditions in the Trump Won treatment group (respectively N=419 and N=381). Error bars reflect 95% confidence intervals. Top horizontal bars show *p*-values for *t* tests of equality of means between different experimental conditions.

Appendix Table B1 provides evidence that individual characteristics are balanced across all four experimental conditions, confirming that the randomization was successful. Figure 1 displays the main findings from this experiments. In the *Clinton won* condition, we observe a large and statistically significant wedge between donation rates in private and in public: a drop from 30% in private to 20% in public (the p-value of a t test of equality is <0.001). Among individuals in the *Trump won* condition, we observe no difference in private and public donation rates, which are 31% and 33%, respectively (p-value=0.594). Moreover, we find no significant difference in private donation rates between the *Clinton won* and *Trump won* conditions (p-value=0.894), suggesting that the information about the winner of the popular vote is not changing privately-held xenophobia. The increase

in public donation rates between the two conditions is statistically significant (p-value<0.001), as is the difference in differences between donation rates in private across conditions and donation rates in public across conditions (p-value=0.008). These results indicate that the information provided about Trump having won the popular vote in the Pittsburgh area causally increased the social acceptability of the action to the point of eliminating the social stigma associated with it among the participants who were told that Clinton won the popular vote. Table 1 displays the difference in differences results in regression format and shows that our results are unchanged when individual covariates are included.[15] The table also displays p-values from permutation tests, showing that our findings are robust to that inference method.

3.3 Evidence of Mechanism

Appendix Figure B2 shows that, consistent with the underlying mechanism of updates in perceptions about the local popularity in xenophobic views, the *Trump won* treatment shifts the distribution of participants' perceptions about the local popularity of these views to the right. The average belief goes up from 42.6% to 50.1% (p-value<0.001).

Given this strong shift in perceptions caused by the treatment, we also use the *Trump won* treatment as an instrument for perceptions. In Appendix Table B3, we present evidence that participants' perceptions causally affect their donation in public, but not in private. A one percentage point increase in the perception of the local popularity of xenophobia increases donation rates in public by 2.3 percentage points.[16]

---

[15]Heterogeneous treatment effects by race, gender, age, marital status, education, and income are reported in Appendix Table B2. The direction of the treatment effect is the same in all subgroups, and differences in the magnitude of the effects between subgroups are never statistically significant. Point estimates indicate, however, that the wedge in donations between public and private in the Clinton won treatment is lower among Whites than among non-Whites, and that the reduction in the wedge in the Public condition is complete among Whites but only partial among non-Whites.

[16]We view the IV results as suggestive, since it is possible that the *Trump won* treatment affects donation rates through other channels. However, we find this unlikely. Indeed, all alternative channels not associated with social acceptability are ruled out by the fact that the donations in private do not change. A remaining possibility is that the *Trump won* treatment might have changed the perceived local acceptability of other behaviors, for example of acting selfishly: after all, Donald Trump can be seen both as the anti-immigrants candidate and as the candidate of greed and self-interest (we thank an anonymous referee for raising this point). The results of Experiment 3 (reported in Appendix F), however, help us rule out this hypothesis. The design of Experiment 3 is similar to that of Experiment 1: it uses donation decisions made either in a private or in a public condition to study the social acceptability of a view. In Experiment 3, however, instead of varying the perceived local popularity of candidate Trump as we do in Experiment 1, we directly randomize the perceived local popularity of anti-Muslim sentiments. Consistently with an update in the perceived popularity of the view being the mechanism at play in both Experiment 1 and 3, we find similar patterns in both experiments.

Table 1: **Experiment 1: Difference in Differences Regressions**

Dependent Variable	Dummy: individual authorizes donation to anti-immigrant organization	
	(1)	(2)
Public	-0.103***	-0.108***
	[0.031]	[0.031]
	(0.002)	(0.000)
Trump Won Treatment	0.004	0.001
	[0.032]	[0.033]
	(0.889)	(0.987)
Public*Trump Won Treatment	0.120***	0.129***
	[0.045]	[0.045]
	(0.005)	(0.001)
Mean Donation Rate Clinton Won Private Treatment	0.304	
Controls	No	Yes
N	1,600	1,587
R^2	0.012	0.023

Notes: Columns (1) presents OLS regression of a dummy variable for whether a individual donates to the anti-immigration organization on a dummy for the public condition, a dummy for the Trump Won condition, and a dummy for the Trump Won public condition. The Clinton Won private condition is the omitted group, for which we report the mean donation rate. Column (2) replicates and add individual covariates (gender, age, marital status, years of education, household income, and race). Robust standard errors in brackets. *P*-values from permutation tests with 1,000 repetitions in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1% based on robust standard errors.

4 Experiment 2: Judging Xenophobic Expression

4.1 Experimental Design

In November 2018, an online survey panel survey company recruited participants (N=1,830) from the U.S. who had previously self-identified as Democrats.[17] Each panelist could participate in the survey only once. The survey was conducted on the online platform Qualtrics.

---

[17]The company we used for this experiment is Prime Panels. See the survey script in Appendix C.3. The reader can access the survey information sheet and interactively follow the survey at `http://ssd.az1.qualtrics.com/jfe/form/SV_4VgnEZSmikzSf8p`.

First, all participants answered a number of demographic questions. The participants of this experiment were randomized into one of four conditions, corresponding to the four treatments of Experiment 1: the *Clinton won, private* group, the *Clinton won, public* group, the *Trump won, private* group, and the *Trump won, public* group. Those in the two *Clinton won* groups were told that we surveyed, in another survey, participants from an area where Hillary Clinton won in 2016. Those in the two *Trump won* groups were told instead that we surveyed participants from an area where Donald Trump won in 2016.

In Experiment 1, we studied how turning off social pressure against xenophobia affected xenophobic expression. In Experiment 2, we focus on how turning on social pressure for xenophobic expression affects judgment and sanctioning of this type of expression. To that end, we made an ex-ante design choice not to disclose to the participants of Experiment 2 that the previous study was about Pittsburgh. This would allow for a larger positive update in perceptions about the local popularity of xenophobia stemming from information about Trump's victory in the area of the first study (and it is indeed what we observe empirically, as shown in Appendix Figures B2 and B3, which display participants' beliefs for both experiments).

All subjects were then presented with two anti-immigrant quotes (the same used in Experiment 1), and were asked to predict the share of voters in the other participant's area that they believed would agree with the quotes. This provids two measures of the beliefs that the participants in this experiment had about the popularity of anti-immigrant sentiments in the area where the previous study took place.

Next, the subjects (players 1) were informed that they had been matched with a participant from the previous survey (player 2). Each player 1 was matched with a random player 2 from one of the four original conditions. For example, a player 1 randomized into the *Clinton won, private* group for Experiment 2 was matched with a player 2 that was randomized into the *Clinton won, private* condition in Experiment 1. Players 1, however, were not informed that the previous survey was an experiment with different treatment conditions. The subjects were then told that player 2 authorized a donation to an anti-immigrant organization, after being shown the exact text of the question in which the donation was authorized. Importantly, the text included either the text of the *private* treatment or of the *public* treatment, so that the subject could fully understand the condition under which the other participant made the donation choice. For example, players 1 in the *Trump won, public* group knew that the player 2 they were matched with was from an area where Donald Trump won the 2016 election and decided to authorize a donation to the anti-immigration organization knowing that their donation decision would be published online, possibly along with their personal details. The subjects, however, were not informed about the fact that the donation in the previous survey was incentivized.

Finally, players 1 were asked to play a dictator game in which they could decide how to split $2 between themselves and player 2. The subject were told that their decision on how much to give

to the other participant was anonymous, and that when making his donation decision, player 2 did not know that they would be playing this follow-up game.

Link to Theory.   This experiment looks at receivers' interpretation of senders' decisions. As above, we interpret type A as xenophobic and type B as tolerant; action A as the senders' xenophobic action (authorizing the donation to the anti-immigration organization) and action B as the tolerant one (not authorizing this donation). Absent social image concerns, xenophobic senders' should have strictly preferred action A (they helped an aligned organization) and tolerant ones should have strictly preferred action B (they refused associating with the organization). Moreover, the subjects of Experiment 2 (players 1 in the dictator game) are the receivers in the model, who judge the action of the senders and decide how much to share with them in a dictator game. All the senders (players 2) have chosen action A, since in Experiment 1 they decided to authorize the donation to the anti-immigration organization: their type, however, is not directly known, but can be inferred by the receivers. We think of participants as type-B members of the audience, since they are all Democrats (and thus unlikely to be xenophobic). Finally, we interpret information that Trump won as the high signal about the share of type A in the senders area, $s = p_H$, and information that Clinton won as the low signal $s = p_L$, for which we believe all type-B tolerant senders would never take the xenophobic action A in public (as in Proposition 3).

The model predicts that in private only type-A senders will chose action A, so, for the receivers, the posterior that player 2 who chose the xenophobic action is a xenophobe equals 1 in the private case, regardless of which candidate won locally. According to Proposition 3, in public, this posterior is also equal to 1 in the case of the low signal $s = p_L$ (there is no social pressure for someone tolerant to act as a xenophobe in the *Clinton won, public* treatment). In contrast, with the high signal $s = p_H$, some tolerant senders might choose the xenophobic action due to social pressure. As a result, the posterior that player 2 who chose the xenophobic action is a xenophobe be would less than 1 in the *Trump won, public* treatment.

Since all receivers (player 1) in our experiment are registered Democrats, the lower posterior that player 2 is a xenophobe should result in lower punishment / higher altruism towards player 1. Thus, we would expect donations by dictators to be higher in the *Trump won, public* treatment than in any other treatment, and the donations in the other three treatments to be similar.[18]

4.2   Main Results

Appendix Table B4 provides evidence that individual characteristics are balanced across all four experimental conditions, confirming that the randomization was successful. Figure 2 displays our main findings from Experiment 2.

---

[18]In Appendix A.5 we provide a simple microfoundation for the link between posteriors about player 1 and donations to player 1 based on homophily and altruism, as discussed in footnote 10.

Figure 2: **Experiment 2: Donation Rates**
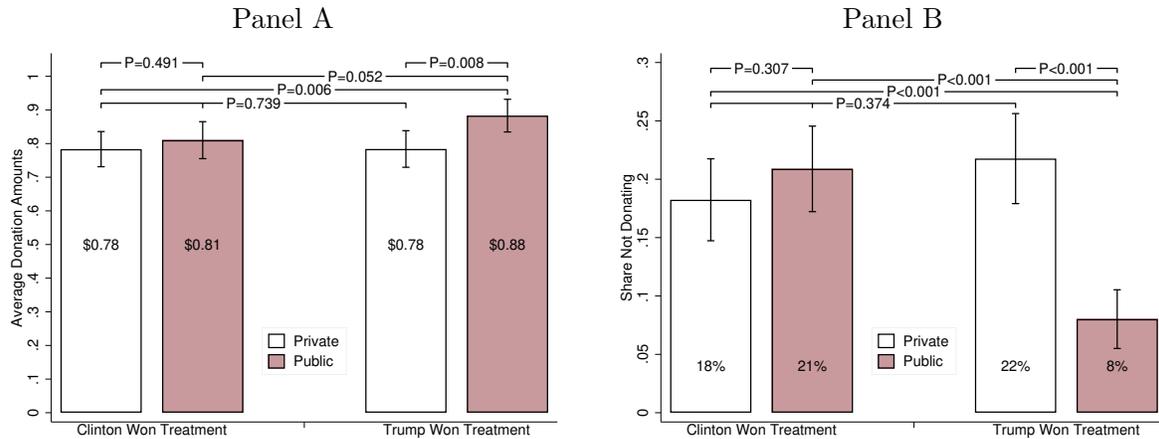
Notes: Panel A displays average donation amounts to the individuals from Experiment 1 in the four experimental conditions. The two bars on the left display donations for individuals in the private and public conditions in the Clinton Won treatment group (respectively N=466 and N=474), and the two bars on the right display those for individuals in the private and public conditions in the Trump Won treatment group (respectively N=441 and N=449). Panel B displays the percent of subjects not making positive donations. Error bars reflect 95% confidence intervals. Top horizontal bars show *p*-values for *t* tests of equality of means between different experimental conditions.

Panel A displays comparisons of average donations across groups. In the *Clinton won, private*, *Clinton won, public*, and *Trump won, private* conditions the average amount given is very similar, at respectively $0.78, $0.81, and $0.78. The p-value of a test of joint equality of the three averages is 0.739. In each of these conditions there was no pressure to donate to the anti-immigrant organization in order to pander to the majority: in the two private conditions there is no social pressure, and in the *Clinton won, public* condition the desire to pander to the majority would, if anything, pressure the subject into not authorizing the donation. From the decision to donate, the subject could then infer that player 2 was likely to truly hold xenophobic views. The lack of difference across the two private conditions helps us deal with alternative interpretations, such as learning about local motives to privately oppose immigration (since shifting perceptions on who won the election in player 2's area could have changed the perception of what that area is).

In the *Trump won, public* condition the average donation is $0.88, higher than in the *Clinton won, private* condition (p-value 0.006), the *Clinton won, public* condition (p-value 0.052), and *Trump won, private* condition (p-value 0.008). The one in the *Trump won, public* condition is the only donation that could have been driven by the desire of player 2 to pander to the majority, complicating the inference that player 1 could make about the anti-immigration views of the

participant from the previous study.

Panel B compares the share of participants who do *not* share anything from their $2 endowment with player 2. Here again, the percentage of subject deciding not to transfer anything to the other participant is similar across the *Clinton won, private*, *Clinton won, public*, and *Trump won, private* conditions, at respectively 18%, 21% and 22%. The p-value of a test of joint equality of the three averages is 0.374. Importantly, the share of participants not donating is 8% in the *Trump won, public* group, significantly lower than in the other three conditions (p-value <0.001 for all three pairwise comparisons).[19]

Appendix Table B5 displays the results in regression format and show that our results are not changed when individual covariates are included.

In Appendix E we present the design and results from a similar experiment conducted on *mTurk*, where participants were asked to play a dictator game with another respondent in Switzerland, and where we manipulated perceptions of the popularity of anti-Muslim sentiment in Switzerland, by randomly giving information about the 2009 Swiss referendum that banned the construction of minarets in that country. The results are again consistent with the predictions of our framework.

4.3 Evidence of Mechanism

Appendix Figure B3 shows that, consistent with the underlying mechanism of updates in perceptions about the local popularity in xenophobic views in the area of player 2, the *Trump won* treatment shifts the distribution of the dictator's perceptions about the local popularity of these views to the right. The average belief goes up from 44.9% to 53.5% (p-value<0.001).

We can again use the *Trump won* treatment as an instrument for perceptions. In Appendix Table B6, we present evidence that the dictators' perceptions causally affect their amount given and the probability of giving when the subject from Experiment 1 donated in public, but not when they donated in private. These results indicate that manipulating beliefs about the popularity of Donald Trump: 1) has the expected effect on the perceived popularity of anti-immigrants views in the area of player 2; 2) this decreases punishment in the public condition (which is when the social pressure for the recipient to express those views could be present); and 3) this does not have any effect in the private condition (where social pressure is absent).

5 Conclusion

In this paper, we study how social norms, usually thought of as relatively stable and persistent, can change rapidly when new information becomes available. In our first experiment, we show that a positive, experimentally-induced update in people's beliefs about Donald Trump's popularity increased their willingness to publicly express xenophobic views. We see no evidence that the

---

[19]The median amount given was $1 in all four treatments, so we do not use it as an outcome.

election increased the likelihood of *having* such views and therefore conclude that the increased expression of certain views should be attributed to a shift of social norms rather than individual preferences or attitudes. Using dictator games, we also test the model's prediction that individuals are judged less negatively for expressing a view that is popular in their environment, and find that it is indeed the case.

Our findings shed light on the factors that can trigger a rapid change in social norms, and in particular, norms against the expression of xenophobic views. Our results suggest that social norms regarding the expression of such views in the U.S. might have been causally changed by Trump's rise in popularity and eventual electoral victory. More broadly, the mechanisms we study in this paper might help explain the rise – and potential consequences – of other crucial recent events such as the *Brexit* vote in the U.K., and more generally the rise in anti-immigrant and anti-minority sentiment in the developed world.

Our analysis suggests at least two lines for subsequent work. One deals with the joint evolution of individual views and social norms. While we see no evidence that Donald Trump's election changed people's views on immigration in the short run, it is well possible that the changed social norm will expose people to views that will eventually influence their own. These individual views could eventually affect both social norms and political decisions. Thus, understanding how individuals acquire and change their preferences through social interactions is of utmost importance. An interesting and important question, for example, is whether laws prohibiting certain speech (such as those banning denial of the Holocaust in Germany and some other countries) are more effective in forming public opinion as compared to cases where such speech is not banned but highly stigmatized (as, e.g., in the U.S.)

A different set of questions stems from our dictator game experiments. We observed that subjects were largely willing to forgive the individual if he publicly expressed xenophobic views as part of conforming to the social norm. Yet they were remarkably unwilling to forgive the individual for holding such views, despite knowing little about the reasons why he acquired them. This alone would be consistent with subjects viewing people from other settings as similar to them as individuals, but living in different social environments, but this explanation is perhaps too simplistic. Nevertheless, understanding how people judge thoughts and actions of people from their own and from different societies and cultures, and perhaps ultimately why social norms emerge, is another interesting avenue for future research.

References

Acemoglu, Daron, and Matthew O. Jackson. 2017. "Social Norms and the Enforcement of Laws." *Journal of the European Economic Association*, 15(2): 245.

Acemoglu, Daron, Tarek A Hassan, and Ahmed Tahoun. 2018. "The Power of the Street: Evidence from Egypt's Arab Spring." *The Review of Financial Studies*, 31(1): 1–42.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *The Quarterly Journal of Economics*, 128(2): 469–530.

Algan, Yann, and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review*, 100(5): 2060–92.

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

Ali, S. Nageed, and Roland Bénabou. 2016. "Image Versus Information: Changing Societal Norms and Optimal Privacy." National Bureau of Economic Research Working Paper 22203.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31(2): 211–36.

Andreoni, James, and B. Douglas Bernheim. 2009. "Social Image and the 5050 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica*, 77(5): 1607–1636.

Andreoni, James, Justin M. Rao, and Hannah Trachtman. 2017. "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving." *Journal of Political Economy*, 125(3): 625–653.

Andreoni, James, Nikos Nikiforakis, and Simon Siegenthaler. 2017. "Social Change and the Conformity Trap." *Mimeo.*

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652–1678.

Bénabou, Roland, and Jean Tirole. 2011. "Laws and Norms." National Bureau of Economic Research Working Paper 17579.

Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841–877.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2018. "Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia." National Bureau of Economic Research Working Paper 24736.

Bursztyn, Leonardo, and Robert Jensen. 2015. "How Does Peer Pressure Affect Educational Investments?" *The Quarterly Journal of Economics*, 130(3): 1329.

Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria*." *The Quarterly Journal of Economics*, 102(2): 179.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–1451.

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

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2017. "Voting to Tell Others." *The Review of Economic Studies*, 84(1): 143.

Elias, Julio J., Nicola Lacetera, and Mario Macis. 2016. "Efficiency-Morality Trade-Offs in Repugnant Transactions: A Choice Experiment." National Bureau of Economic Research Working Paper 22632.

Enikolopov, Ruben, Alexey Makarin, Maria Petrova, and Leonid Polishchuk. 2017. "Social Image, Networks, and Protest Participation." *Mimeo*.

Enke, Benjamin. 2020. "Moral Values and Voting." *Journal of Political Economy*, Forthcoming.

Fernández, Raquel. 2007. "Women, Work, and Culture." *Journal of the European Economic Association*, 5(2-3): 305–332.

Fowler, James H., and Cindy D. Kam. 2007. "Beyond the Self: Social Identity, Altruism, and Political Participation." *Journal of Politics*, 69(3): 813–827.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102(1): 3348.

Giani, Marco, and Pierre-Guillaume Méon. 2019. "Global Racist Contagion Following Donald Trump's Election." *British Journal of Political Science*, 1–8.

Giuliano, Paola. 2007. "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association*, 5(5): 927–952.

Greif, Avner. 2006. "History Lessons: The Birth of Impersonal Exchange: The Community Responsibility System and Impartial Justice." *Journal of Economic Perspectives*, 20(2): 221–236.

Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, 86(3): 653–60.

Janus, Alexander L. 2010. "The Influence of Social Desirability Pressures on Expressed Immigration Attitudes*." *Social Science Quarterly*, 91(4): 928–946.

Katz, Daniel, and Floyd H. Allport. 1931. *Students' Attitudes: A Report of the Syracuse University Reaction Study.* Syracuse, NY:Craftsman Press.

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying Social Norms using Coordination Games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Kuran, Timur. 1991. "The East European Revolution of 1989: Is It Surprising That We Were Surprised?" *American Economic Review*, 81(2): 121–125.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification.* Cambridge, MA:Harvard University Press.

Lohmann, Susanne. 1993. "A Signaling Model of Informative and Manipulative Political Action." *The American Political Science Review*, 87(2): 319–333.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan Weigel. 2017. "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom." *Econometrica*, 85(4): 1065–1091.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do Political Protests Matter? Evidence from the Tea Party Movement." *The Quarterly Journal of Economics*, 128(4): 1633.

Milgrom, Paul, and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica*, 62(1): 157–180.

Morris, Stephen. 2001. "Political Correctness." *Journal of Political Economy*, 109(2): 231–265.

Müller, Karsten, and Carlo Schwarz. 2019. "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment."

O'Gorman, Hubert J. 1975. "Pluralistic Ignorance and White Estimates of White Support for Racial Segregation." *Public Opinion Quarterly*, 39(3): 313.

Perez-Truglia, Ricardo, and Guillermo Cruces. 2017. "Partisan Interactions: Evidence from a Field Experiment in the United States." *Journal of Political Economy*, 125(4): 1208–1243.

Prendergast, Canice. 1993. "A Theory of 'Yes Men'." *American Economic Review*, 83(4): 757–770.

Richman, Barak D. 2017. "An Autopsy of Cooperation: Diamond Dealers and the Limits of Trust-Based Exchange." *Journal of Legal Analysis*, 9(2): 247–283.

Rushton, J. Philippe. 1989. "Genetic similarity, human altruism, and group selection." *Behavioral and Brain Sciences*, 12(3): 503518.

Voigtländer, Nico, and Hans-Joachim Voth. 2012. "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *The Quarterly Journal of Economics*, 127(3): 1339–1392.

Xiong, Heyu. 2017. "Media Personality and Its Political Premium." *Mimeo.*

Supplementary Appendix
(Not For Publication)

A Theory Appendix

A.1 Proofs

Proof of Proposition 1.

In the private case, $n = 0$, and thus each citizen chooses $d_i = t_i$. So consider the public case. Slightly abusing notation, let us write $U_{A,h}(d_i)$ to denote the utility $U_i(d_i)$ of citizen i with $t_i = A$ and $h_i = h$, and let us write $U_{B,h}(d_j)$ to denote the utility $U_j(d_j)$ of citizen j with $t_j = B$ and $h_j = h$.

In what follows, we denote the equilibrium shares of types A and types B who choose action A by r_A and r_B, respectively. Then an individual with prior belief P about the share of type A (where $P = P_A$ for type A and $P = P_B$ for type B, where P_A and P_B are defined in the text, believes that a decision-maker who chose $d_i = A$ has type A with probability $Q_{P,A} = \frac{Pr_A}{Pr_A + (1-P)r_B}$, while one who chose $d_i = B$ has type A with probability $Q_{P,B} = \frac{P(1-r_A)}{P(1-r_A)+(1-P)(1-r_B)}$.

Our first observation is that the following increasing differences property holds in any equilibrium: $U_{A,h}(A) - U_{A,h}(B) > U_{B,h}(A) - U_{B,h}(B)$. Suppose, to obtain a contradiction, than this is not the case. Let us write down these conditions:

$$V_A + h\left(P_A Q_{P_A,A} + (1-P_A)(1-Q_{P_B,A})\right) - h\left(P_A Q_{P_A,B} + (1-P_A)(1-Q_{P_B,B})\right) \leq$$
$$h\left(P_B Q_{P_A,A} + (1-P_B)(1-Q_{P_B,A})\right) - V_B - h\left(P_B Q_{P_A,B} + (1-P_B)(1-Q_{P_B,B})\right).$$

Rearranging, we get

$$V_A + V_B + h\left(P_A - P_B\right)\left(Q_{P_A,A} + Q_{P_B,A} - Q_{P_A,B} - Q_{P_B,B}\right) \leq 0.$$

The latter bracket cannot exceed 2, since all the terms are probabilities lying on $[0, 1]$. Thus, we have

$$V_A + V_B \leq 2h\left(P_A - P_B\right).$$

Since $p_L < P_A < P_B < p_H$, we have $P_A - P_B < p_H - p_L$. However, by Assumption 1, $V_A, V_B \geq h(p_L - p_H)$. This contradiction implies the increasing differences property.

This increasing differences property implies that any equilibrium that satisfies the D1 criterion has the feature that if all citizens choose $d_i = B$ on equilibrium path, then anyone who deviates to choose action A will be considered to be type A (and the other way around). Indeed, it means that

the benefit from deviating to action A, $U_i(A) - U_i(B)$, cannot be maximized for an individual of type B.

To proceed, notice that if $U_{A,h}(A) \geq U_{A,h}(B)$ for some h, then $U_{A,h'}(A) > U_{A,h'}(B)$ for $h' < h$. Indeed, suppose not. Denoting

$$L = P_A Q_{P_A,A} + (1 - P_A)(1 - Q_{P_B,A}) - P_A Q_{P_A,B} - (1 - P_A)(1 - Q_{P_B,B}),$$

we have $V_A + hL \geq 0$ and $V_A + h'L \leq 0$, which implies $(h - h')L \geq 0$, so $L \geq 0$. However, then $V_A + h'L > 0$, a contradiction. Similarly if $U_{B,h}(B) \geq U_{B,h}(A)$ for some h, then $U_{B,h'}(B) \geq U_{B,h'}(A)$ for $h' < h$. This implies that A types may choose action B only if their social image concerns exceed a certain cutoff, and the same applies to B types choosing action A. More precisely, if r_A is the share of A types choosing action A, then all those with $h < F(r_A)$ choose action A and all those with $h > F(r_A)$ choose action B (either set may be empty); similarly, B types with $h < 1 - F(r_B)$ choose action B and those with $h > 1 - F(r_B)$ choose action A.

Let us show that either $r_A = 1$ or $r_B = 0$. Suppose not, so $r_A < 1$ and $r_B > 0$. This means that citizen i with $t_i = A$ and $h_i = M$ chooses B and citizen j with $t_j = B$ and $h_j = M$ chooses A. This is only possible if $U_{A,M}(A) \leq U_{A,M}(B)$ and $U_{B,M}(A) \geq U_{B,M}(B)$. However, this contradicts the single crossing property, so either $r_A < 1$ or $r_B = 0$ in any equilibrium.

Now fix $r_B = 0$ and suppose that h is the cutoff social image concern for type A, so $r_A = F(h)$. Consider $G_A(h) = U_{A,h}(A \mid r_A = F(h), r_B = 0) - U_{A,h}(B \mid r_A = F(h), r_B = 0)$ as a function of h; let us show that it has at most one zero, and if so it changes the sign from positive to negative. We have

$$\begin{aligned}
G_A(h) &= V_A + h(P_A Q_{P_A,A} + (1 - P_A)(1 - Q_{P_B,A})) \\
&\quad - h(P_A Q_{P_A,B} + (1 - P_A)(1 - Q_{P_B,B})) \\
&= V_A + h\left(P_A - P_A \frac{P_A(1 - F(h))}{1 - P_A F(h)} - (1 - P_A)\left(1 - \frac{P_B(1 - F(h))}{1 - P_B F(h)}\right)\right) \\
&= V_A + h(1 - P_A)\left(\frac{P_A}{1 - P_A F(h)} - \frac{1 - P_B}{1 - P_B F(h)}\right).
\end{aligned}$$

Suppose that $G_A(h) = 0$ for some h. This implies, in particular, that $\frac{P_A}{1 - P_A F(h)} < \frac{1 - P_B}{1 - P_B F(h)}$. It suffices to prove that in that case $\frac{dG_A(h)}{dh} < 0$. We have

$$\begin{aligned}
\frac{d_A G(h)}{dh} &= hf(h)(1 - P_A)\left(\frac{(P_A)^2}{(1 - P_A F(h))^2} - \frac{P_B(1 - P_B)}{(1 - P_B F(h))^2}\right) \\
&\quad + (1 - P_A)\left(\frac{P_A}{1 - P_A F(h)} - \frac{1 - P_B}{1 - P_B F(h)}\right).
\end{aligned}$$

Consider two cases. If $P_B \geq \frac{1}{2}$, then $\frac{P_B}{1-P_B} \geq 1$, so

$$\frac{P_B(1-P_B)}{(1-P_BF(h))^2} = \frac{P_B}{1-P_B}\frac{(1-P_B)^2}{(1-P_BF(h))^2} \geq \left(\frac{1-P_B}{1-P_BF(h)}\right)^2 > \left(\frac{P_A}{1-P_AF(h)}\right)^2,$$

which means that both terms in $\frac{dG_A(h)}{dh}$ are negative, which proves that $\frac{dG_A(h)}{dh} < 0$. So suppose $P_B < \frac{1}{2}$, which means $\frac{P_B}{1-P_B} < 1$. We have

$$\begin{aligned}
\frac{1}{1-P_A}\frac{dG_A(h)}{dh} &= hf(h)\left(\frac{(P_A)^2}{(1-P_AF(h))^2} - \frac{(1-P_B)^2}{(1-P_BF(h))^2}\right) + \left(\frac{P_A}{1-P_AF(h)} - \frac{1-P_B}{1-P_BF(h)}\right)\\
&\quad + hf(h)\frac{1-2P_B}{1-P_B}\frac{(1-P_B)^2}{(1-P_BF(h))^2}\\
&= \left(\frac{P_A}{1-P_AF(h)} - \frac{1-P_B}{1-P_BF(h)}\right)\left(hf(h)\left(\frac{P_A}{1-P_AF(h)} + \frac{1-P_B}{1-P_BF(h)}\right) + 1\right)\\
&\quad + hf(h)\frac{1-2P_B}{1-P_B}\frac{(1-P_B)^2}{(1-P_BF(h))^2}.
\end{aligned}$$

After plugging in $\frac{P_A}{1-P_AF(h)} - \frac{1-P_B}{1-P_BF(h)} = -\frac{V_A}{h(1-P_A)}$, we get

$$\begin{aligned}
\frac{1}{1-P_A}\frac{dG_A(h)}{dh} &= -\frac{V_A}{h(1-P_A)}\left(hf(h)\left(\frac{P_A}{1-P_AF(h)} + \frac{1-P_B}{1-P_BF(h)}\right) + 1\right)\\
&\quad + hf(h)\frac{1-2P_B}{1-P_B}\frac{(1-P_B)^2}{(1-P_BF(h))^2}\\
&\leq -\frac{V_A}{h(1-P_A)}\left(hf(h)(P_A + 1 - P_B) + 1\right) + hf(h)\frac{1-2P_B}{1-P_B}\\
&< -\frac{V_A}{h}\left(hf(h) + 1\right) + hf(h)\\
&= \frac{-V_A + hf(h)(h-V_A)}{h} \leq \frac{-V_A + V_A}{h} = 0,
\end{aligned}$$

proving the result.

We can similarly prove that if we fix $r_A = 1$ and consider $G_B(h) = U_{B,h}(A \mid r_A = 1, r_B = 1 - F(h)) - U_{B,h}(B \mid r_A = 1, r_B = 1 - F(h))$, which may be written as

$$\begin{aligned}
G_B(h) &= h(P_BQ_{P_A,A} + (1-P_B)(1-Q_{P_B,A}))\\
&\quad - V_B - h(P_BQ_{P_A,B} + (1-P_B)(1-Q_{P_B,B}))\\
&= h\left(P_B\frac{P_A}{1-(1-P_A)F(h)} + (1-P_B)\left(1 - \frac{P_B}{1-(1-P_B)F(h)}\right) - (1-P_B)\right) - V_B\\
&= hP_B\left(\frac{P_A}{1-(1-P_A)F(h)} - \frac{1-P_B}{1-(1-P_B)F(h)}\right) - V_B,
\end{aligned}$$

27

then it has at most one zero, and if so, it changes the sign from negative to positive.

Now consider the following cases. If $V_A \geq M(1 - 2P_A)$ and $V_B \geq M(2P_B - 1)$, then $G_A(M) \geq 0$ and $G_B(M) \leq 0$. This implies that there is an equilibrium where all A types choose A and all B types choose B, and since $G_A(h) > 0$ and $G_A(h) < 0$ for all $h < M$, this is the only equilibrium. Notice that conditions $V_A < M(1 - 2P_A)$ and $V_B < M(2P_B - 1)$ are mutually exclusive (adding the right-hand sides yields $2(P_B - P_A) < 0$). If $V_A < M(1 - 2P_A)$ and $V_B \geq M(2P_B - 1)$, then all B types choose action B, as do some A types, and in that case all A types choose B if and only if $G_A(m) \leq 0$. Similarly, if $V_A \geq M(1 - 2P_A)$ and $V_B < M(2P_B - 1)$, then all citizens choose A if and only if $G_B(m) \geq 0$. This gives the equilibrium characterization.

Let us prove the comparative statics with respect to θ. Notice that both P_A and P_B are increasing as a function of θ. This means that an increase in θ may switch the equilibrium type from (v) to (iv) to (iii) etc, but not the other way around. Now suppose that for a given θ the equilibrium type is (iv), so all B types choose action B, and the cutoff h for A types is such that $G_A(h) = 0$. Consider a marginal increase in θ; since $\frac{dG_A(h)}{dh} < 0$, it suffices to prove that $G_A(h)$ is strictly increasing in θ at such h. For this, it is sufficient to prove that $G_A(h)$ has a positive derivative with respect to P_A and nonnegative with respect to P_B whenever $G_A(h) = 0$. We have

$$
\begin{aligned}
\frac{\partial G_A(h)}{\partial P_A} &= h\left(\frac{1 - P_B}{1 - P_B F(h)} - \frac{P_A}{1 - P_A F(h)}\right) + h\frac{1 - P_A}{1 - P_A F(h)} \\
&= \frac{V_A}{1 - P_A} + h\frac{1 - P_A}{1 - P_A F(h)} > 0; \\
\frac{\partial G_A(h)}{\partial P_B} &= h(1 - P_A)\frac{1 - F(h)}{(1 - P_B F(h))^2} \geq 0.
\end{aligned}
$$

This proves that the share of type A choosing action A is strictly increasing in θ if the equilibrium type is (iv). The case of equilibrium type (ii) is considered similarly. This completes the proof. ∎

Proof of Proposition 2.

The posterior probability that $p = p_H$ conditional on signal s is given by $\theta(s)$, which is defined in the main text. The posteriors of citizens of either type who, in addition to the signal, condition on their realized type, are given by the same expressions as P_A and P_B with θ replaced by $\theta(s)$. The equilibrium in this case given by Proposition 1 with θ being replaced by $\theta(s)$.

Notice that for $\mu > \frac{1}{2}$, $\theta(p_L) < \theta < \theta(p_H)$. The result follows immediately from the comparative statics with respect to θ. This completes the proof. ∎

Proof of Proposition 3.

Suppose that for $s = p_L$, citizens with type $t_i = B$ choose action $d_i = B$. Then $\Pr_j(t_i = A \mid d_i = A, t_j, s = p_L) = 1$ for either type of citizen $j \neq i$; this follows from Bayes formula if $G_A(M) > 0$ in equilibrium or from that the equilibrium satisfies the D1 criterion (see the proof of Proposition 1) if $G_A(M) = 0$.

Then the corresponding probability for signal $s = p_H$, $\Pr_j (t_i = A \mid d_i = A, t_j, s = p_H) \leq 1 = \Pr_j (t_i = A \mid d_i = A, t_j, s = p_L)$. Furthermore, if for $s = p_H$, $G_B (M) > 0$, so some citizens with type $t_i = B$ choose action $d_i = A$, then $\Pr_j (t_i = A \mid d_i = A, t_j, s = p_H) < 1$. This completes the proof. ∎

A.2 Multiple equilibria

If we do not impose Assumption 1, there may be multiple equilibia. One reason for multiplicity is given by the following example.

Example 1. *Let $V_A = V_B = 1$, $p_H = \frac{99}{100}$, $p_L = \frac{2}{3}$, $\theta = \frac{9}{10}$. Let $m = M = 2$, so h has a degenerate distribution (this may be easily generalized to a smooth distribution with high density). In this way, the first part of Assumption 1 is violated, while the second is satisfied.*

Here, we have $P_A \approx 0.97$ and $P_B \approx 0.74$. One can verify that there are three equilibria. In one, every citizen chooses action A; in another, both types of citizens choose their preferred action, and there is also a third equilibrium where all A types choose action A, while B types split, with about 48% choosing action A and the rest choosing B.

The intuition behind this multiplicity is the following. Choosing one's preferred action is an equilibrium, because an actor of type B believes that there is a sufficient share of type B citizens in the audience. Importantly, if everyone chooses their preferred action, then this action allows the audience to infer the actor's type precisely. Now suppose we make some B actors switch into choosing action A. In this case, if a B type citizen chooses action B, then everyone will still believe he is B, while choosing action A is inconclusive. However, the posterior of a type A audience member will go down very little (her prior that a citizen is type A is very strong), while the posterior of a type B audience member will be more sensitive, and she will put a substantial probability on the actor being type B. This effect makes choosing action A more attractive. In other words, if there is some pooling in equilibrium, then for an individual decision-maker, pooling allows to hide true identity and allow every member of the audience to make the inference she finds more likely, which likely corresponds to her type.

The next example shows that the second part of Assumption 1 is also necessary to ensure equilibrium uniqueness.

Example 2. *Suppose $V_A = V_B = 1$, $p_H = 0.9$, $p_L = 0.1$, $\theta = \frac{1}{2}$. Then type A citizens believe that others are type A with probability $P_A = 0.82$, while type B citizens believe that others are type A with probability $P_B = 0.18$. Let h be distributed on $[m, M]$ such that $m \geq \frac{25}{16} = 1.5625$ (an example of a distribution that satisfies this and also the first requirement of 1 is one with p.d.f. $f(h) = \frac{1}{h(h-1)}$ for $h \in \left[\frac{25}{16}, \frac{25}{25-9e} \right]$ and $f(h) = 0$ otherwise).*

In this example, there are two equilibria. The first has all A citizens choosing $d_i = A$ and all B citizens choosing $d_i = B$. Indeed, then $r_A = 1$ and $r_B = 0$, and the expected utilities of type A from the two actions are

$$
\begin{aligned}
U_{A,h}(A) &= V_A + h\left(P_A \frac{P_A r_A}{P_A r_A + (1 - P_A) r_B} + (1 - P_A)\left(1 - \frac{P_B r_A}{P_B r_A + (1 - P_B) r_B}\right)\right) \\
&= V_A + h P_A; \\
U_{A,h}(B) &= h\left(P_A \frac{P_A(1 - r_A)}{P_A(1 - r_A) + (1 - P_A)(1 - r_B)} + (1 - P_A)\left(1 - \frac{P_B(1 - r_A)}{P_B(1 - r_A) + (1 - P_B)(1 - r_B)}\right)\right) \\
&= h(1 - P_A),
\end{aligned}
$$

and we thus have

$$
U_{A,h}(A) - U_{A,h}(B) = V_A + h P_A - h(1 - P_A) = 1 + 0.64h > 0.
$$

This means that citizens of type A do not want to deviate, and, similarly, neither do citizens of type B.

However, there is another equilibrium, where all A citizens choose $d_i = B$ and all B citizens choose $d_i = A$. Indeed, if they do, then $r_A = 0$ and $r_B = 1$. If so, the expected utilities of type A from the two actions are

$$
\begin{aligned}
U_{A,h}(A) &= V_A + h(1 - P_A); \\
U_{A,h}(B) &= h P_A.
\end{aligned}
$$

We thus have

$$
U_{A,h}(A) - U_{A,h}(B) = V_A + h(1 - P_A) - h P_A = 1 - 0.64h \leq 1 - 0.64 \times \frac{25}{16} = 0.
$$

This means that type A citizens are indeed better off choosing action B (and almost all of them strictly so), and likewise B citizens prefer action A. Thus, it is an equilibrium.

The intuition behind Example 2 is simple. Citizens of different types may have very different beliefs about the distribution of types in the society, and they overestimate the share of citizens of their own type relative to other players. The example shows that it is possible that everyone chooses the "opposite" action for signaling purposes (and the signals are interpreted correctly, i.e., action A is interpreted as evidence that the actor is of type B). Since both types believe that their type is common, deviating is not profitable, because the audience of their own type will fail to recognize them as one of their own. Despite this simple intuition, we believe that this equilibrium is hardly realistic, and imposed the second part of Assumption 1 to ensure that citizens care about the action enough to rule it out, so V_A and V_B are not too small relative to social image concerns.

Even though we impose Assumption 1 in the model, the comparative statics results would hold more generally. For example, if the first part of Assumption 1 is violated, the signs of the derivatives with respect to θ (as computed in the proof of Proposition 1) are still the same. This implies that even if there are multiple equilibria, the comparative statics results hold for all stable equilibria (with stability appropriately defined), in particular the largest and the smaller ones (in the sense of the share of citizens choosing, e.g., action A).

A.3   Numerical example

Figure A1: **Numerical Example**
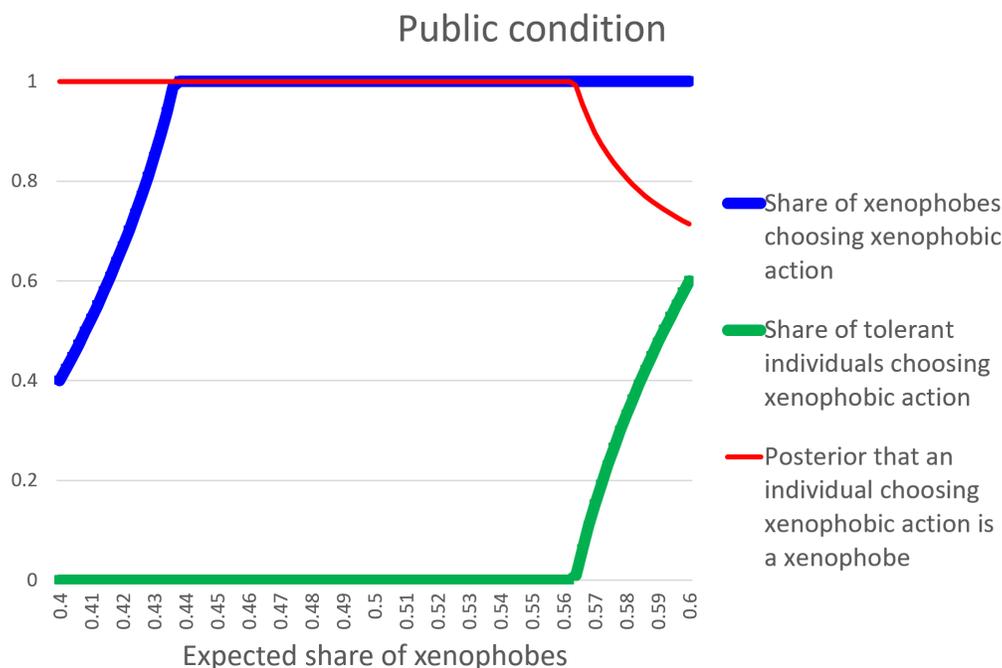
We illustrate the comparative statics with a simple example. As before, we interpret type A and action A as xenophobic and type B and action B as tolerant. The society may be either "more tolerant" with share of A types $p_L = 0.4$ or "more xenophobic" with this share equal to $p_H = 0.6$. We will allow parameters θ (the prior that the society has $p = p_H$), μ (signal precision), and the realization of signal s to vary. By varying these parameters, the posterior probability that $p = p_H$, $\theta(s)$, can take any value between 0 and 1, and so the posterior expectation of the share of xenophobic types A can be anywhere between $p_L = 0.4$ and $p_H = 0.6$. Normalize the utility levels of both types to choose their preferred action to $V_A = V_B = 1$ and assume that social image concerns h_i are distributed uniformly on $[0, 5]$. For these parameter values, both parts of Assumption 1 are

satisfied, and therefore there is a unique equilibrium.

We will allow parameters θ (the prior that the society has $p = p_H$), μ (signal precision) and the realization of signal s to vary. By varying these parameters, the posterior expectation of the share of xenophobic types A can be anywhere between $p_L = 0.4$ and $p_H = 0.6$. On Figure A1, we plot this posterior expectation on the horizontal axis, and we plot the shares of A and B types choosing action A, as well as the posterior that someone who chose action A is an A type. We do so in the public case; in the private case, each type would choose their preferred action, and the posterior that someone choosing an action is that type would equal 1.

As one can see, in the public case, tolerant types choose the tolerant action, except if they are sufficiently sure that there are many xenophobic types. Similarly, xenophobic types choose the xenophobic action, unless they believe that tolerant types are prevalent. These effects are symmetric given the symmetry of parameter values. There is also a range of posterior beliefs where all individuals choose their preferred actions. Note that these graphs are monotone: a higher posterior about the share of xenophobic individuals increases the propensity of individuals of either type to choose the xenophobic action, and sometimes strictly so. This illustrates Proposition 2.

On the receivers' side, an increase in the posterior about the share of xenophobic individuals has a nonmonotonic effect. If everyone chooses their preferred action, or if some xenophobes choose tolerant action because of social pressure to do so, anyone who is observed choosing a xenophobic action must be a xenophobe. This is not the case where the posterior about the share of xenophobes is high enough: there, some tolerant individuals choose xenophobic action, and the posterior probability that someone who chose a xenophobic action is truly a xenophobe becomes less than 1. This illustrates Proposition 3.

Note that the assumption made in Proposition 3 is important: the graph of the posterior that an individual who chose action A is indeed type A does not have to be monotonically decreasing more generally. For example, if $p_H = 1$, then this graph would increase back to 1 as the society becomes sufficiently convinced that almost all individuals are xenophobes. In other words, higher social pressure pushes this curve down, while the direct effect of a higher expected number of xenophobes pushes it up.

A.4   Private signals

Throughout the model, we consider signal s to be public. This is motivated by our experimental design. For example, in the experiment during the 2016 election, our subjects presumably knew that if they are contacted by a researcher, this researcher would know the information about Trump's local popularity that was provided to them. In Experiment 1, because the subjects were not aware of the other treatments, it is likely that they thought that other subjects would be provided the same information, so they would know about Trump's or Clinton's local popularity when accessing the website that we provided. All this points to public signals, where the actor expects the audience

to have gotten the signal as well, as the natural assumption in the light of our experiments.

Nevertheless, it is possible that some subjects did not understand or believe that the audience will likely have the same information. For example, in the Experiment 1, even though the link to the survey results was provided to subjects only, some might have thought that this information would be available more broadly (perhaps they thought that their friends would find the results by googling their names). This would correspond to a private signal, where the actor gets the signal about the value of p, but the audience does not. In other words, the actor would then maximize

$$U_i\left(d_i; s_{priv}\right) = V_{t_i}\mathbf{I}\left\{d_i = t_i\right\} + \mathbf{I}_{pub}h_i \sum_{t\in\{A,B\}} \Pr_i\left(t_j = t \mid t_i, s_priv\right)\Pr_j\left(t_i = t \mid d_i, t_j = t\right);$$

the difference is that the latter probability is not conditional on the signal s_{priv} because the audience does not receive the signal.

It is straightforward to show that the comparative statics predictions of the model (Propositions 2 and 3) continue to hold in the case of private signals. The easiest way to see that is to consider the functions $G_A\left(h\right)$ and $G_B\left(h\right)$ defined in the proof of Proposition 1 and show that they are monotonically increasing in direct inclusions of P_A and P_B (which correspond to the actor's beliefs of distribution of A types in the audience, conditional on his type) while holding the inclusions through $Q_{P_A,A}, Q_{P_A,B}, Q_{P_B,A}, Q_{P_B,B}$ (which correspond to audience member's posterior conditional on her type and the actor's decision) fixed. This immediately yields the comparative statics result as in Proposition 2, whereas the proof of Proposition 3 remains literally the same. In other words, the predictions of the model do not qualitatively depend on whether the signals are private or public, i.e. if the decision-makers are aware that the audience has the same information or not.

A.5 Model of receiver's action

In the model, we assumed that the sender gets direct benefit if the receiver believes they are similar. The implicit reason was that this has an unspecified and unmodeled future benefit, which in practice could be a higher probability of becoming friends, getting hired, or having a more efficient communication. In the experiment, however, some participants of Experiment 1 were in fact rewarded in the dictator game by participants of Experiment 2 in a way consistent with higher donations if participants of Experiments 2 thought of them as similar. Here we build a simple model of the dictator game where the decision-maker cares about similarity.

Suppose that the decision-maker divides a budget B between himself and another participant. Each player's utility is $u\left(c\right)$, which is increasing, concave, and for simplicity satisfies the Inada conditions. The decision-maker cares about the other participant at some rate $r \in (0, 1)$. We can think of rate r as the sum of several effects. The may be a baseline altruism a, capturing the decision-maker's minimum care for any other individual in the society. On top of that, the weight he puts on the utility of the other individual may be higher if he believes that the two individuals

are similar. This effect has been documented in the literature on sociobiology for genetic similarities (e.g., Rushton, 1989), and in the context of dictator games (dividing a budget) for social distance (Hoffman, McCabe and Smith, 1996) and political distance (Fowler and Kam, 2007). Let us thus write $r = a + f(s)$, where f is an increasing function of s, which is a measure of similarity between the two agents as perceived by the decision-maker.

The decision-maker therefore solves the problem

$$\max u(c_1) + r u(c_2)$$

$$\text{s.t. } c_1 \geq 0, c_2 \geq 0, c_1 + c_2 \leq B.$$

In the solution to this problem, the amount the decision-maker gives to the other player, c_2, satisfies

$$u'(B - c_2) - r u'(c_2) = 0.$$

Since the left-hand side is increasing in c_2 (because $u(\cdot)$ is concave) and decreasing in r, then c_2 must be increasing in r. This implies, in particular, that a higher degree of similarity s would lead to a higher transfer by the decision-maker.

B   Appendix Figures and Tables

Figure B2: **Experiment 1: Beliefs About Others**
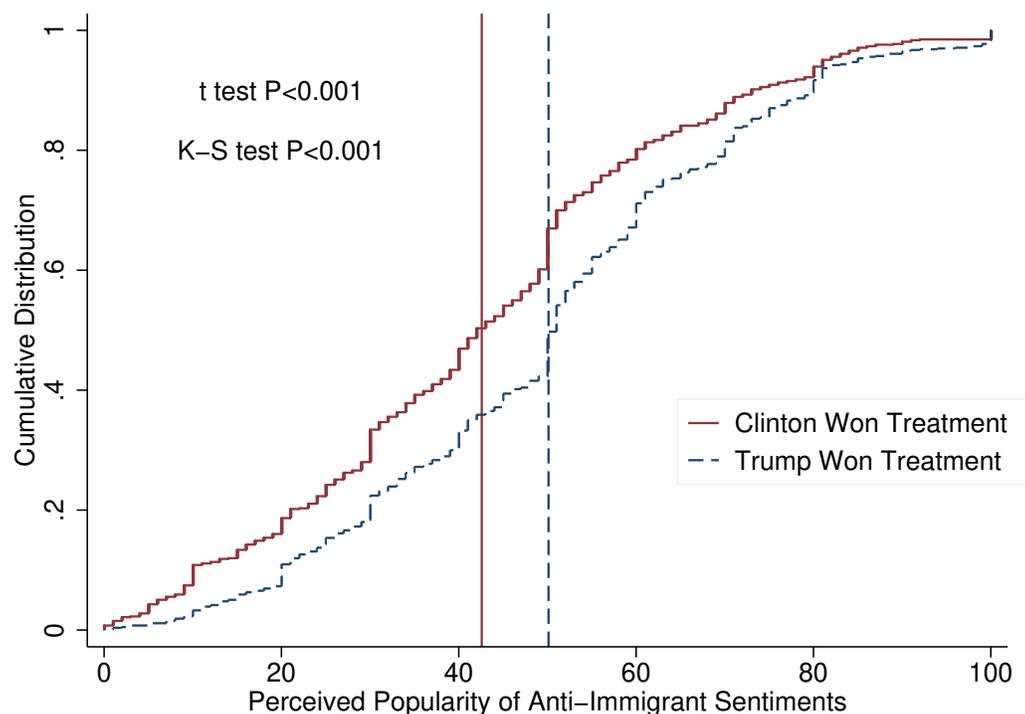
Notes: Empirical cumulative distributions of perceived popularity of anti-immigrant sentiments for individuals in the Clinton Won treatment and in the Trump Won treatment (respectively N=793 and N=794). The two vertical lines display the means of the two distributions. K-S P is the *p*-value of a Kolmogorov-Smirnov test of equality of the two distributions, while *t* test P is the *p*-value of a test of equality of means.

Figure B3: **Experiment 2: Beliefs About Others**
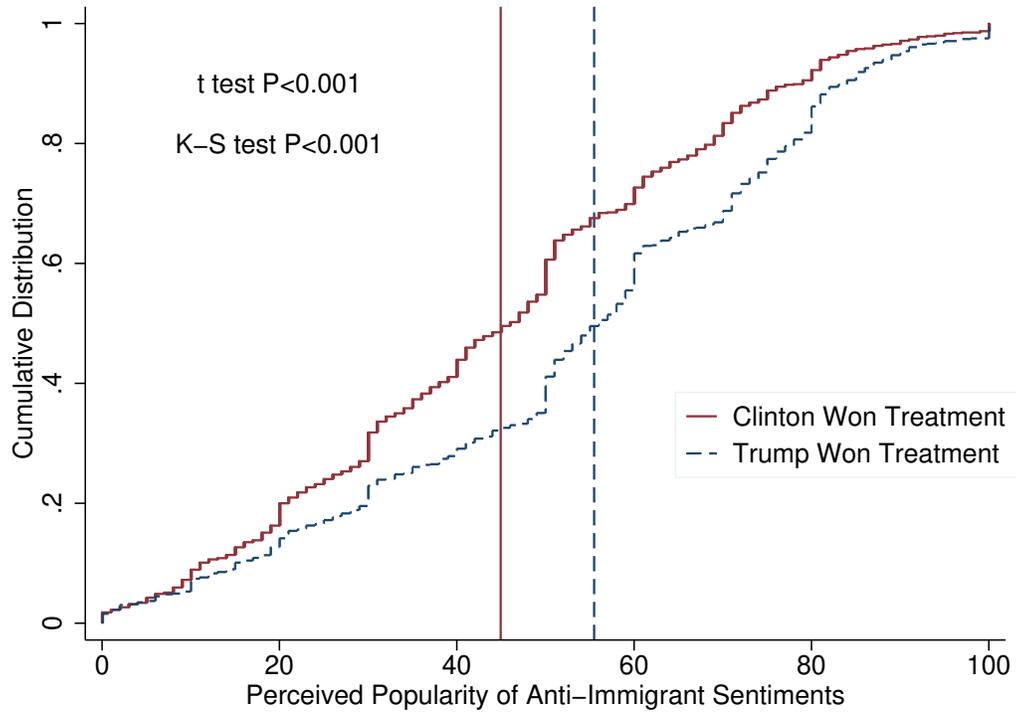
Notes: Empirical cumulative distributions of perceived popularity of anti-immigrant sentiments for individuals in the Clinton Won treatment and in the Trump Won treatment (respectively N=940 and N=890). The two vertical lines display the means of the two distributions. K-S P is the *p*-value of a Kolmogorov-Smirnov test of equality of the two distributions, while *t* test P is the *p*-value of a test of equality of means.

Table B1: **Experiment 1: Balance of Covariates**

	Full Sample	Clinton Won Private	Clinton Won Public	Trump Won Private	Trump Won Public	p-*value*
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.49	0.50	0.49	0.50	0.46	0.711
	[0.500]	[0.501]	[0.501]	[0.501]	[0.499]	
Age	42.89	42.64	43.91	42.60	42.35	0.480
	[15.231]	[14.784]	[15.522]	[14.955]	[15.669]	
Married	0.44	0.44	0.44	0.45	0.44	0.997
	[0.497]	[0.497]	[0.498]	[0.498]	[0.497]	
Education	14.19	14.22	14.27	14.09	14.17	0.680
	[2.220]	[2.203]	[2.283]	[2.093]	[2.307]	
Household Income	56323.25	55556.99	56941.03	57103.37	55582.01	0.909
	[39331.754]	[39317.398]	[39758.092]	[40007.064]	[38248.176]	
White	0.90	0.89	0.91	0.92	0.89	0.512
	[0.296]	[0.312]	[0.284]	[0.278]	[0.311]	
No Demographics	0.01	0.02	0.00	0.01	0.01	0.215
	[0.090]	[0.123]	[0.050]	[0.084]	[0.089]	
Totals	1600	392	408	419	381	

Notes: Column (1) reports the mean level of each variable, with standard deviations in brackets, for the full sample. Columns (2) to (5) report the mean level of each variable, with standard deviations in brackets, for all the experimental conditions. Column (6) reports the p-value of a test that means are the same in all the experimental conditions.

Table B2: Experiment 1: Heterogeneity

	White		Female		Age		Married		Education		Income	
	No	Yes	No	Yes	Below Median	Above Median	No	Yes	Below Median	Above Median	Below Median	Above Median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Public	-0.194**	-0.095***	-0.121***	-0.087**	-0.128***	-0.083*	-0.094**	-0.117**	-0.134***	-0.075*	-0.091**	-0.120**
	[0.094]	[0.032]	[0.044]	[0.043]	[0.041]	[0.046]	[0.041]	[0.046]	[0.043]	[0.043]	[0.040]	[0.047]
Trump Won Treatment	-0.019	0.005	0.007	-0.002	0.016	-0.012	-0.010	0.018	-0.043	0.048	0.008	-0.008
	[0.108]	[0.034]	[0.047]	[0.045]	[0.044]	[0.048]	[0.043]	[0.050]	[0.046]	[0.046]	[0.043]	[0.049]
Public*Trump Won Treatment	0.075	0.131***	0.134**	0.110*	0.101*	0.160**	0.160***	0.079	0.196***	0.051	0.130**	0.118*
	[0.138]	[0.048]	[0.064]	[0.063]	[0.060]	[0.068]	[0.060]	[0.068]	[0.063]	[0.065]	[0.060]	[0.068]
Mean Donation Rate	0.333	0.299	0.330	0.276	0.287	0.321	0.293	0.316	0.312	0.295	0.270	0.343
Clinton Private Treatment												
N	154	1,433	811	776	838	749	883	704	796	791	851	736
R^2	0.035	0.013	0.015	0.009	0.017	0.013	0.014	0.014	0.018	0.011	0.014	0.012

Notes: This table reports heterogeneous treatment effects. Each column shows results from a separate OLS regression of a dummy variable for whether an individual donates to the anti-immigration organization on a dummy for the public condition, a dummy for the Trump Won condition, and a dummy for the Trump Won public condition. The Clinton Won private condition is the omitted group, for which we report the mean donation rate. Each column shows the results for a different subsample across six traits. "Age" is a dummy that is equal to one for individuals above the median age in the sample, "Education" is a dummy that is equal to one for individuals above the median years of education in the sample, and "Income" is a dummy equal to one for individuals above the median income in the sample. Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1% based on robust standard errors.

38

Table B3: **Experiment 1: Instrumental Variable Regressions**

Panel A: First Stage Regressions						
Dependent Variable	Perceived Share of Voters Holding Xenophobic Views					
	(1)	(2)	(3)	(4)	(5)	(6)
Trump Won Treatment	7.528***	7.559***	9.492***	9.385***	5.541***	5.632***
	(1.111)	(1.101)	(1.568)	(1.547)	(1.580)	(1.573)
Mean of Dependent Variable	42.588		41.451		43.666	
in the Clinton Won Treatment						
Panel B: Instrumental Variable Regressions						
Dependent	Dummy: individual authorizes donation to anti-immigrant organization					
	(1)	(2)	(3)	(4)	(5)	(6)
Perceived Share of Voters			0.000	0.000	0.023***	0.024***
Holding Xenophobic Views			(0.003)	(0.003)	(0.007)	(0.008)
Mean Donation Rate						
at a 50% Perceived Share of Voters			30.507		34.350	
Holding Xenophobic Views						
Sample	Full Sample		Private Treatment		Public Treatment	
Controls	No	Yes	No	Yes	No	Yes
N	1587	1587	802	802	785	785

Notes: Panel A presents OLS regressions of the perceived share of voters holding xenophobic views on a dummy for the Trump Won treatment. Specifically, we measure the percentage of Pittsburgh voters the respondents believed would agree with the quote "For European American society and culture to persist requires a European-American majority, and a clear one at that." The Clinton Won treatment is the omitted group, for which we report the mean perceived share. For columns (3), (4), (5), and (6) these estimates also represent the first stage of the instrumental variable regressions presented in Panel B. Panel B presents IV regressions of a dummy variable for whether an individual donates to the anti-immigration organization on the perceived share of voters holding xenophobic views. The latter is instrumented with the Trump Won treatment. We substract 50 from the measure of the share, so that the incercept of the regression represents the average donation rate at a perceived share of 50%. Columns (2) and (4) and (6) replicate and add individual covariates (gender, age, marital status, years of education, household income, and race). Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1% based on robust standard errors.

Table B4: **Experiment 2: Balance of Covariates**

	Full Sample	Clinton Won Private	Clinton Won Public	Trump Won Private	Trump Won Public	p-*value*
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.51	0.52	0.49	0.53	0.51	0.586
	[0.500]	[0.500]	[0.500]	[0.500]	[0.500]	
Age	44.50	45.59	43.55	44.41	44.47	0.329
	[17.054]	[17.298]	[16.424]	[16.820]	[17.664]	
Married	0.39	0.37	0.38	0.40	0.40	0.683
	[0.487]	[0.482]	[0.485]	[0.491]	[0.490]	
Education	14.30	14.24	14.39	14.32	14.26	0.747
	[2.270]	[2.361]	[2.261]	[2.148]	[2.306]	
Household Income	54961.75	55568.67	55738.40	55170.07	53307.35	0.747
	[37560.939]	[38580.298]	[37503.814]	[37076.771]	[37093.439]	
White	0.77	0.76	0.76	0.80	0.79	0.357
	[0.418]	[0.426]	[0.430]	[0.402]	[0.410]	
Totals	1830	466	474	441	449	

Notes: Column (1) reports the mean level of each variable, with standard deviations in brackets, for the full sample. Columns (2) to (5) report the mean level of each variable, with standard deviations in brackets, for all the experimental conditions. Column (6) reports the *p*-value of a test that means are the same in all the experimental conditions.

Table B5: **Experiment 2: Regressions**

Dependent Variable	Average donation		Dummy: no donation	
	(1)	(2)	(3)	(4)
Clinton Won Private group	-0.099***	-0.099***	0.102***	0.101***
	[0.036]	[0.037]	[0.022]	[0.022]
	(0.008)	(0.009)	(0.000)	(0.000)
Clinton Won Public group	-0.073*	-0.072*	0.129***	0.127***
	[0.037]	[0.037]	[0.023]	[0.023]
	(0.044)	(0.048)	(0.000)	(0.000)
Trump Won Private group	-0.099***	-0.100***	0.138***	0.137***
	[0.037]	[0.037]	[0.023]	[0.023]
	(0.008)	(0.007)	(0.000)	(0.000)
Trump Won Public group	0.883		0.080	
Controls	No	Yes	No	Yes
N		1830		
R^2	0.005	0.011	0.021	0.035

Notes: Columns (1) presents an OLS regression of the donation amount to the previous survey participant on a dummy for the Clinton Won private group, a dummy for the Clinton Won public group, and a dummy for the Trump Won private group. Trump Won public group is the omitted group, for which we report the mean donation amount. Columns (3) presents an OLS regression of a dummy variable for subjects not making positive donations to the other survey participant on treatment dummies. Trump Won public group is the omitted group, for which we report the share of subjects not making positive donations. Columns (2) and (4) replicate and add individual covariates (gender, age, marital status, years of education, household income, and race). Robust standard errors in brackets. *P*-values from permutation tests with 1,000 repetitions in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1% based on robust standard errors.

Table B6: **Experiment 2: Instrumental Variable Regressions**

Panel A: First Stage Regressions						
Dependent Variable	Perceived Share of Voters Holding Xenophobic Views					
	(1)	(2)	(3)	(4)	(5)	(6)
Trump Won Treatment	8.544***	8.779***	6.438***	6.619***	10.614***	10.896***
	(1.155)	(1.148)	(1.666)	(1.661)	(1.560)	(1.587)
Mean of Dependent Variable	44.948		45.118		44.781	
in the Clinton Won Treatment						
Panel B: Instrumental Variable Regressions						
Dependent	Amount Donated in the Dictator Game					
	(1)	(2)	(3)	(4)	(5)	(6)
Perceived Share of Voters			0.000	0.000	0.007*	0.006*
Holding Xenophobic Views			(0.006)	(0.006)	(0.004)	(0.003)
Mean Donation Rate						
at a 50% Perceived Share of Voters			78.380		84.600	
Holding Xenophobic Views						
Panel C: Instrumental Variable Regressions						
Dependent	Dummy: individual does not share anything in the dictator game					
	(1)	(2)	(3)	(4)	(5)	(6)
Perceived Share of Voters			0.005	0.005	-0.012***	-0.012***
Holding Xenophobic Views			(0.005)	(0.004)	(0.003)	(0.003)
Mean Donation Rate						
at a 50% Perceived Share of Voters			20.916		14.558	
Holding Xenophobic Views						
Sample	Full Sample		Private Treatment		Public Treatment	
Controls	No	Yes	No	Yes	No	Yes
N	1830	1830	907	907	923	923

Notes: Panel A presents OLS regressions of the perceived share of voters holding xenophobic views on a dummy for the Trump Won treatment. Specifically, we measure the percentage of Pittsburgh voters the respondents believed would agree with the quote "For European American society and culture to persist requires a European-American majority, and a clear one at that." The Clinton Won treatment is the omitted group, for which we report the mean perceived share. For columns (3), (4), (5), and (6) these estimates also represent the first stage of the instrumental variable regressions presented in Panel B. Panel B presents IV regressions of a dummy variable for whether an individual donates to the anti-immigration organization on the perceived share of voters holding xenophobic views. The latter is instrumented with the *Trump* treatment. We substract 50 from the measure of the share, so that the incercept of the regression represents the average donation rate at a perceived share of 50%. Columns (2) and (4) and (6) replicate and add individual covariates (gender, age, marital status, years of education, household income, and race). Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1% based on robust standard errors.

C Survey Scripts

C.1 Demographics

- What is your state of legal residence?
- What is your gender?
  - Male
  - Female
- What is your year of birth?
- What is your marital status?
  - Single
  - Married
- How would you describe your ethnicity/race? Please, check all that apply.
  - White or European American
  - Black or African American
  - Hispanic or Latino
  - Asian or Asian American
  - Other
- What is the highest level of school you have completed or the highest degree you have received?
  - Less than high school degree
  - High school graduate (high school diploma or equivalent including GED)
  - Some college but no degree
  - Associate degree in college (2-year)
  - Bachelor's degree in college (4-year)
  - Master's degree
  - Doctoral degree
  - Professional degree (JD, MD)

- What is your household annual income? Please indicate the answer that includes your entire household income in 2015 before taxes.

  - Less than $10,000
  - $10,000 to $19,999
  - $20,000 to $29,999
  - $30,000 to $39,999
  - $40,000 to $49,999
  - $50,000 to $59,999
  - $60,000 to $69,999
  - $70,000 to $79,999
  - $80,000 to $89,999
  - $90,000 to $99,999
  - $100,000 to $149,999
  - $150,000 or more

C.2 Experiment 1

Experiment 1: Trump Won Private

- This study focuses on political and social beliefs in Pittsburgh. Below are three interesting facts about Pittsburgh politics. Which of these facts did you know before reading them here?

  - Before Pittsburgh was chartered as a city in 1816, it elected a Chief Burgess rather than a mayor. The first Chief Burgess of Pittsburgh, George Robinson, was arrested on orders from George Washington due to his involvement in the Whiskey Rebellion.

  - Pittsburgh native and Olympic qualifying boxer Art Rooney was instrumental in the founding of the Pittsburgh Steelers in the 1930s and the Pittsburgh Penguins in the 1960s. His grandson, Thomas Rooney, serves in the House of Representatives.

  - In the 2016 US Presidential Election, Donald Trump won Pittsburgh's metropolitan area.

- We will now randomly select one among two different organizations, and will give you the opportunity to authorize a donation to the selected organization:
  One is an organization which seeks to reduce overall migration to the United States;
  One is an organization which welcomes immigrants to the United States.
  If you choose to authorize the donation, you will receive an additional $1, on top of your base payment for completing the survey. The $1 donation will not be subtracted from your base payment, so your total compensation will be your base payment plus the extra $1.
  If instead you choose not to donate, you will only receive your base payment for completing the survey.

- The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR).**
  The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion" who wrote "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."
  The anonymized results from this survey will be posted on our website in approximately one month. Results will be reported as percentages only, so your individual response to the survey will remain anonymous. We will notify you when the results become available on our website, [website link].
  Would you like us to donate $1 to FAIR on your behalf?

  - Yes, please donate $1 to FAIR on my behalf.

  - No, please do not donate $ to FAIR on my behalf.

- We will now present two quotes. For each, we will ask you to report what percentage of Pittsburgh voters you believe would agree with the quote.

- "For European American society and culture to persist requires a European-American majority, and a clear one at that."

- "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

Experiment 1: Clinton Won Private

- This study focuses on political and social beliefs in Pittsburgh. Below are three interesting facts about Pittsburgh politics. Which of these facts did you know before reading them here?

  - Before Pittsburgh was chartered as a city in 1816, it elected a Chief Burgess rather than a mayor. The first Chief Burgess of Pittsburgh, George Robinson, was arrested on orders from George Washington due to his involvement in the Whiskey Rebellion.

  - Pittsburgh native and Olympic qualifying boxer Art Rooney was instrumental in the founding of the Pittsburgh Steelers in the 1930s and the Pittsburgh Penguins in the 1960s. His grandson, Thomas Rooney, serves in the House of Representatives.

  - In the 2016 US Presidential Election, Hillary Clinton won Pittsburgh's county.

- We will now randomly select one among two different organizations, and will give you the opportunity to authorize a donation to the selected organization:
  One is an organization which seeks to reduce overall migration to the United States;
  One is an organization which welcomes immigrants to the United States.
  If you choose to authorize the donation, you will receive an additional $1, on top of your base payment for completing the survey. The $1 donation will not be subtracted from your base payment, so your total compensation will be your base payment plus the extra $1.
  If instead you choose not to donate, you will only receive your base payment for completing the survey.

- The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR).**
  The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion" who wrote "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."
  The anonymized results from this survey will be posted on our website in approximately one month. Results will be reported as percentages only, so your individual response to the survey will remain anonymous. We will notify you when the results become available on our website, [website link].
  Would you like us to donate $1 to FAIR on your behalf?

  - Yes, please donate $1 to FAIR on my behalf.

  - No, please do not donate $ to FAIR on my behalf.

- We will now present two quotes. For each, we will ask you to report what percentage of Pittsburgh voters you believe would agree with the quote.

- "For European American society and culture to persist requires a European-American majority, and a clear one at that."

- "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

Experiment 1: Trump Won Public

- This study focuses on political and social beliefs in Pittsburgh. Below are three interesting facts about Pittsburgh politics. Which of these facts did you know before reading them here?

  - Before Pittsburgh was chartered as a city in 1816, it elected a Chief Burgess rather than a mayor. The first Chief Burgess of Pittsburgh, George Robinson, was arrested on orders from George Washington due to his involvement in the Whiskey Rebellion.

  - Pittsburgh native and Olympic qualifying boxer Art Rooney was instrumental in the founding of the Pittsburgh Steelers in the 1930s and the Pittsburgh Penguins in the 1960s. His grandson, Thomas Rooney, serves in the House of Representatives.

  - In the 2016 US Presidential Election, Donald Trump won Pittsburgh's metropolitan area.

- We will now randomly select one among two different organizations, and will give you the opportunity to authorize a donation to the selected organization:
  One is an organization which seeks to reduce overall migration to the United States;
  One is an organization which welcomes immigrants to the United States.
  If you choose to authorize the donation, you will receive an additional $1, on top of your base payment for completing the survey. The $1 donation will not be subtracted from your base payment, so your total compensation will be your base payment plus the extra $1.
  If instead you choose not to donate, you will only receive your base payment for completing the survey.

- The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR).**
  The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion" who wrote "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."
  The results from this survey, including your individual donation decision and the donation decisions of all of the other Pittsburgh respondents to this survey, will be posted on our website in approximately one month. There is no need to provide your name, email, etc. here; the survey company we work with has this information already. We will notify you when the results become available on our website, [website link].
  Would you like us to donate $1 to FAIR on your behalf?

  - Yes, please donate $1 to FAIR on my behalf.

  - No, please do not donate $ to FAIR on my behalf.

- We will now present two quotes. For each, we will ask you to report what percentage of Pittsburgh voters you believe would agree with the quote.

- "For European American society and culture to persist requires a European-American majority, and a clear one at that."

- "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

Experiment 1: Clinton Won Public

- This study focuses on political and social beliefs in Pittsburgh. Below are three interesting facts about Pittsburgh politics. Which of these facts did you know before reading them here?

    - Before Pittsburgh was chartered as a city in 1816, it elected a Chief Burgess rather than a mayor. The first Chief Burgess of Pittsburgh, George Robinson, was arrested on orders from George Washington due to his involvement in the Whiskey Rebellion.

    - Pittsburgh native and Olympic qualifying boxer Art Rooney was instrumental in the founding of the Pittsburgh Steelers in the 1930s and the Pittsburgh Penguins in the 1960s. His grandson, Thomas Rooney, serves in the House of Representatives.

    - In the 2016 US Presidential Election, Hillary Clinton won Pittsburgh's county.

- We will now randomly select one among two different organizations, and will give you the opportunity to authorize a donation to the selected organization:
One is an organization which seeks to reduce overall migration to the United States;
One is an organization which welcomes immigrants to the United States.
If you choose to authorize the donation, you will receive an additional $1, on top of your base payment for completing the survey. The $1 donation will not be subtracted from your base payment, so your total compensation will be your base payment plus the extra $1.
If instead you choose not to donate, you will only receive your base payment for completing the survey.

- The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR).**
The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion" who wrote "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."
The results from this survey, including your individual donation decision and the donation decisions of all of the other Pittsburgh respondents to this survey, will be posted on our website in approximately one month. There is no need to provide your name, email, etc. here; the survey company we work with has this information already. We will notify you when the results become available on our website, [website link].
Would you like us to donate $1 to FAIR on your behalf?

    - Yes, please donate $1 to FAIR on my behalf.

    - No, please do not donate $ to FAIR on my behalf.

- We will now present two quotes. For each, we will ask you to report what percentage of Pittsburgh voters you believe would agree with the quote.

- "For European American society and culture to persist requires a European-American majority, and a clear one at that."

- "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

C.3 Experiment 2

Experiment 2: Trump Won Private

- In another survey, we surveyed participatns in an area that **Donald Trump** won in 2016.

- We will now present two quotes. For each, we will ask you to report what percentage of the voters **in the other participant's area** you believe would agree with the quote.
  "For European American society and culture to persist requires a European-American majority, and a clear one at that."
  "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

- You have been matched with a participant from the previous survey. The other participant saw the question in the next screen. **Please read it carefully, since it will affect the amount of your bonus payment.**

- The other participant saw the following question:

  *The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR).***

  The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion," who wrote: "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."

  *The anonymized results from this survey will be posted on our website [later this month]. Results will be reported as percentages only, **so your individual response to the survey will remain anonymous**. We will notify you when the results become available on our website, [website link]. Would you like us to donate $1 to FAIR on your behalf?*

- The other participant authorized us to donate $1 to FAIR on his behalf.

- You and the other participant will split a total bonus of $2. You alone will make the decision of how much of the $2 you will receive and how much of the $2 the other participant will receive. You can choose to divide the $2 however you like, and you get to keep whatever you do not give to the other participant. For example, if you decide to give $1.30, then you will receive $0.70. Your survey provider will credit this bonus to your account.

  How much would you like to give to the other participant? Keep in mind that when making his donation decision, **the other participant did not know that he would be playing this follow-up game**. Your decision about how much to give the other participant will be completely anonymous.

Experiment 2: Clinton Won Private

- In another survey, we surveyed participatns in an area that **Hillary Clinton** won in 2016.

- We will now present two quotes. For each, we will ask you to report what percentage of the voters **in the other participant's area** you believe would agree with the quote.
  "For European American society and culture to persist requires a European-American majority, and a clear one at that."
  "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

- You have been matched with a participant from the previous survey. The other participant saw the question in the next screen. **Please read it carefully, since it will affect the amount of your bonus payment.**

- The other participant saw the following question:

  *The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR)**.*

  The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion," who wrote: "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."

  *The anonymized results from this survey will be posted on our website [later this month]. Results will be reported as percentages only, **so your individual response to the survey will remain anonymous**. We will notify you when the results become available on our website, [website link]. Would you like us to donate $1 to FAIR on your behalf?*

- The other participant authorized us to donate $1 to FAIR on his behalf.

- You and the other participant will split a total bonus of $2. You alone will make the decision of how much of the $2 you will receive and how much of the $2 the other participant will receive. You can choose to divide the $2 however you like, and you get to keep whatever you do not give to the other participant. For example, if you decide to give $1.30, then you will receive $0.70. Your survey provider will credit this bonus to your account.

  How much would you like to give to the other participant? Keep in mind that when making his donation decision, **the other participant did not know that he would be playing this follow-up game**. Your decision about how much to give the other participant will be completely anonymous.

Experiment 2: Trump Won Public

- In another survey, we surveyed participatns in an area that **Donald Trump** won in 2016.

- We will now present two quotes. For each, we will ask you to report what percentage of the voters **in the other participant's area** you believe would agree with the quote.
  "For European American society and culture to persist requires a European-American majority, and a clear one at that."
  "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

- You have been matched with a participant from the previous survey. The other participant saw the question in the next screen. **Please read it carefully, since it will affect the amount of your bonus payment.**

- The other participant saw the following question:

  *The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR)**.*

  The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion," who wrote: "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."

  *The results from this survey, including your individual donation decision and the donation decisions of all of the [other respondents to this survey from your area], will be posted on our website [later this month]. **There is no need to provide your name, email, etc. hear; the survey company we work with has this information already.** We will notify you when the results become available on our website, [website link]. Would you like us to donate $1 to FAIR on your behalf?*

- The other participant authorized us to donate $1 to FAIR on his behalf.

- You and the other participant will split a total bonus of $2. You alone will make the decision of how much of the $2 you will receive and how much of the $2 the other participant will receive. You can choose to divide the $2 however you like, and you get to keep whatever you do not give to the other participant. For example, if you decide to give $1.30, then you will receive $0.70. Your survey provider will credit this bonus to your account.

  How much would you like to give to the other participant? Keep in mind that when making his donation decision, **the other participant did not know that he would be playing this follow-up game**. Your decision about how much to give the other participant will be completely anonymous.

51

Experiment 2: Clinton Won Public

- In another survey, we surveyed participatns in an area that **Hillary Clinton** won in 2016.

- We will now present two quotes. For each, we will ask you to report what percentage of the voters **in the other participant's area** you believe would agree with the quote.
  "For European American society and culture to persist requires a European-American majority, and a clear one at that."
  "Both legal and illegal immigration should be drastically reduced because immigrants undermine American culture and do not respect American values."

- You have been matched with a participant from the previous survey. The other participant saw the question in the next screen. **Please read it carefully, since it will affect the amount of your bonus payment.**

- The other participant saw the following question:

  *The organization randomly chosen for you is the **Federation for American Immigration Reform (FAIR).***

  The Federation for American Immigration Reform is an immigration-reduction organization of concerned individuals who believe that immigration laws must be reformed and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of "The Immigration Invasion," who wrote: "I've come to the point of view that for European American society and culture to persist requires a European-American majority, and a clear one at that."

  *The results from this survey, including your individual donation decision and the donation decisions of all of the [other respondents to this survey from your area], will be posted on our website [later this month]. **There is no need to provide your name, email, etc. hear; the survey company we work with has this information already.** We will notify you when the results become available on our website, [website link]. Would you like us to donate $1 to FAIR on your behalf?*

- The other participant authorized us to donate $1 to FAIR on his behalf.

- You and the other participant will split a total bonus of $2. You alone will make the decision of how much of the $2 you will receive and how much of the $2 the other participant will receive. You can choose to divide the $2 however you like, and you get to keep whatever you do not give to the other participant. For example, if you decide to give $1.30, then you will receive $0.70. Your survey provider will credit this bonus to your account.

  How much would you like to give to the other participant? Keep in mind that when making his donation decision, **the other participant did not know that he would be playing this follow-up game**. Your decision about how much to give the other participant will be completely anonymous.

D Experiment 1B: Expressing Xenophobia in 2016

We implemented a first version of Experiment 1 in the two weeks before and in the week after the 2016 U.S. presidential election. The timing of the experiment allowed us to exploit the uniqueness of the situation and study the process of information aggregation as it was unfolding. We conducted both waves with workers from the online platform *mTurk*. The platform draws workers from very diverse backgrounds, though it is not representative of the U.S. population as a whole.

D.1 Experimental Design

Wave 1: Intervention Before the Election. During the two weeks prior to the presidential election, we recruited participants (N = 458) from the eight states in which the expected probability of Donald Trump's victory at the state level was 100%, according to the website *Predictwise*: Alabama, Arkansas, Idaho, Nebraska, Oklahoma, Mississippi, West Virginia, and Wyoming. *mTurk* workers with at least 80% approval rate could see our request, which was described as a "5 minute survey" with a reward of $0.50. Each worker could participate in the survey only once. Workers who clicked on the request were displayed detailed instructions about the task, and given access to links to the study information sheet and the actual survey. The survey was conducted on the online platform *Qualtrics*.[20]

After answering a number of demographic questions, half of the participants were randomly informed about the 100% local odds from the website (*information* condition) while the other half were not informed (*control* condition). Though restricting to these states might affect the external validity of the findings, it also allows us not to worry about the role of heterogeneous priors (and updates) in response to an informational treatment: the 100% forecast ensured that for this half of the sample, the direction of the update about Trump's local popularity is either zero or positive, but never negative.[21]

Our main goal is to measure the perceived social acceptability of strong anti-immigrant sentiment using a donation experiment with real stakes. Participants were first told that they would be given the opportunity to make a donation to a randomly drawn organization that could either be anti- or pro-immigration, to ensure that participants would not associate the experimenters with a

---

[20]The experiment can be found in the AEA RCT Registry (AEARCTR-0001752). The reader can access the survey information sheet and interactively follow the survey at `http://ssd.az1.qualtrics.com/jfe/form/SV_0Uhwmq3Icp7XSa9`.

[21]In the information condition, before being given the information about the 100% local odds from the website, respondents were asked about their beliefs on the probability of Trump winning the election in their state. This question was asked to respondents in the control condition only at the end of the survey instead. Eliciting priors in the control group to assess the direction of the update would have been challenging since the forecast information was available online. Therefore, asking the question before the donation decision could have undone the treatment. Answers to the question if asked after the donation decision could have been affected by the decision itself and by the private/public condition later assigned to the participant. For analogous reasons, the control condition did not include a placebo statement about Trump, since this could have primed participants to think about his chances of winning the election in the next couple of weeks.

specific political view. To maximize power and avoid direct deception, the randomization was such that more than 90% of participants (N=428) would get assigned the organization we were interested in: the *Federation of American Immigration Reform*.[22] To make sure that the participants were aware of the organization's very strong anti-immigration stance, a few more details about the organization and its founder were provided in the experiment:

> The Federation for American Immigration Reform (FAIR) is an **immigration-reduction organization** of concerned individuals who believe that immigration laws must be reformed, and seeks to reduce overall immigration (both legal and illegal) into the United States. The founder of FAIR is John Tanton, author of 'The Immigration Invasion' who wrote "I've come to the point of view that for European-American society and culture to persist requires a European-American majority, and a clear one at that."

Participants were then asked if they would like to authorize the researchers to donate $1 to that organization on their behalf. The money would not come from the subject's $0.50 payment for participation in the study. Moreover, the participant would also be paid an *extra* $1 (or about 1/6 of an hourly wage on *mTurk*) if he authorized the donation. Rejecting the donation would not affect the monetary payoffs to the participant in any way other than through the loss of this extra amount.

In addition to the original randomization of informing subjects about Trump's probability of victory in the participant's state, we introduced a second layer of cross-randomization at the donation stage. Half of the participants were assured that their donation authorization would be kept completely anonymous, and that no one, not even the researchers would be able to match their decision to their name: we refer to this condition as the *private* condition. Specifically, participants were told:

> Note: just like any other answer to this survey, also **your donation decision will be completely anonymous**. No one, not even the researchers, will be able to match your decision to your name.

The other half of the subjects were instead informed, right before the donation question was displayed to them, that they might be personally contacted by the research team to verify their answers to the questions in the remaining part of the survey: this is what we refer to as the *public* condition.

> Important: in order to ensure the quality of the data collected, a member of the research team **might personally contact you** to verify your answers to the next question and the following ones.

---

[22]The pro-immigration organization was the *National Immigration Forum*. Of course, it is still possible that to some degree the participants assigned to the Federation of American Immigration Reform were more likely to associate the researcher with anti-immigration views than those assigned to the pro-immigration organization.

Names and contact information were not collected during the intervention, since the practice is not allowed on *mTurk*. As a result, it was not possible to credibly lead participants to believe that their decision would be observed by other individuals, for example, from their state. However, on *mTurk* it is possible to contact participants individually on the platform via their worker ID. We were therefore able to minimize deception since the decision was anonymous yet researchers could still potentially contact participants (moreover, participants in the public condition might have believed that they would be asked for personal information in case they were contacted later on).

Wave 2: Intervention After the Election. We exploited the natural experiment of Trump's unexpected victory as an alternative "treatment" that could lead to an increase in the social acceptability of holding xenophobic views. We repeated the experimental intervention in the same states during the first week after the election, restricting the design to the control condition with no additional initial information on Trump's popularity. We analyze the behavior of subjects who were part of our sample before the election (N = 168; 166 of them assigned to the anti-immigration organization). Based on naturally occurring variation, we can assess the impact of Trump's electoral victory on the perceived social acceptability of xenophobia.

D.2   Results

The first four bars of Figure D1 display our main findings from the pre-election experiment. In the control condition before the election, we observe a large and statistically significant wedge between donation rates in private and in public: a drop from 54% in private to 34% in public (the p-value of a t test of equality is 0.002). Among individuals in the information condition, we observe no difference in private and public donation rates, which are 47% and 46%, respectively (p-value=0.839). Moreover, we find no significant difference in private donation rates between the information and control conditions (p-value=0.280), suggesting that the information is not increasing privately-held xenophobia. The increase in public donation rates between the two conditions is statistically significant (p-value=0.089), as is the difference in differences between donation rates in private across conditions and donation rates in public across conditions (p-value=0.050). These results indicate that the information provided causally increased the social acceptability of the action to the point of eliminating the original social stigma associated with it.[23]

As an additional way of examining the effect of Trump's increased popularity on public expression of xenophobia, we compare the private and public donation rates in the control condition

---

[23]Apart from social stigma, another possible reason for the lower donation rates in the public condition with respect to the private condition is that participants might want to avoid talking with the surveyor because of the extra effort and time this requires (independently of the topic of the conversation), and they might expect the likelihood of having to talk to be higher in case they decide to make the donation. However, this mechanism should operate both in the control and in the treatment conditions, thus not affecting our identification of the reduction in social stigma.

before and after the election. In the last two bars of Figure D1, we analyze the actions of respondents who participated in both waves of the experiment. In private, we again observe no increase in donation rates (54% before the election and 49% after the election, p-value=0.440). In public, we observe a significant increase from 34% before the election to 48% after it (p-value=0.060). The difference in differences between donation rates in private before and after the election and donation rates in public before and after the election in the control condition is also statistically significant (p-value=0.062). It is worth emphasizing that the donation rates following the two different "treatments" (either experimental or natural) are extremely similar: 47% vs. 49% in private, and 46% vs. 48% in public.

Figure D1: **Experiment 1B: Donation Rates Before and After the Election**
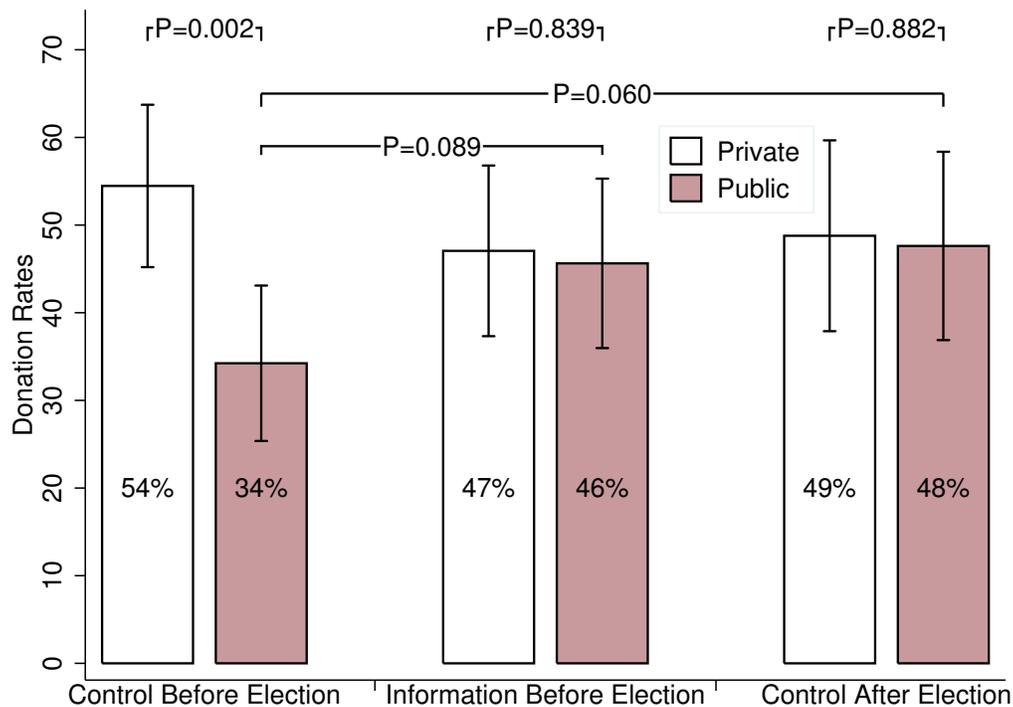
Notes: the two bars on the left display donation rates to the anti-immigration organization for individuals in the private and public conditions in the control group before the election (full sample, respectively N=112 and N=111), the two central bars display those in the information group before the election (full sample, respectively N=102 and N=103), and the last two bars display those in the control group after the election (for individuals already surveyed before the election, respectively N=82 and N=84). Error bars reflect 95% confidence intervals. Top horizontal bars show p-values for t tests of equality of means between different experimental conditions.

E Experiment 2B: Dictator Game with Swiss Player

E.1 Experimental Design

Wave 1: Non-Anonymous Behavior by the Swiss Player In late February 2017, we recruited participants from the six states in which Hillary Clinton won the presidential election with the highest margin: California, Hawaii, Maryland, Massachusetts, New York, and Vermont. This was done to maximize the chances of recruiting subjects with liberal views, and in particular subjects with no anti-Muslim sentiment.[24]

First, after answering a number of demographic questions, all participants were told that a minaret is a tower typically built adjacent to a mosque and traditionally used for the Muslim call to prayer. Second, they were asked whether they would support the introduction of a law prohibiting the building of minarets in their state. Following our pre-registration, we focus on subjects who reported to be against the introduction of this law (N = 396), and we examine how they would interact with a person who has opposite views.[25] In order to do so, in the third part of the survey, participants were told that they were matched with a subject from another survey and were asked to play a dictator game in which they could decide how to split $3 (half of an hourly wage on the platform) between themselves and the other participant. We randomly assigned our participants to three different groups and randomized the background information we gave to our participants about the person they were matched with. Participants in the control group were only told that the participant they were matched with was a 24-year-old male from Switzerland. Note that we used real 24-year-old male subjects from Switzerland recruited to take part in a short survey by a research assistant from the University of Zurich.

Participants in the *anti-minarets* group were additionally told that this person supports the prohibition of the building of minarets in Switzerland. Participants in the *anti-minarets, public support* group were instead told that "like 57.5% of Swiss respondents, the participant supports the prohibition of the building of minarets in Switzerland."

Wave 2: Anonymous Behavior by the Swiss Player If we find higher donations in the *anti-minarets, public support* group, when compared to the *anti-minarets* one, we can conclude that the participants may believe that the Swiss person has strategic reasons to state that he is anti-minarets, and for this reason judge him less for expressing that view. However, a potential

---

[24]As in Experiment 1B, MTurk workers with at least 80% approval rate could see our request, which in this case was described as a "4-5 minutes short survey" with a reward of $0.50. Each worker could participate in the survey only once. Workers who clicked on the request were displayed detailed instructions about the task, and given access to links to the study information sheet and the actual survey. The survey was conducted on the online platform *Qualtrics*. The experiment can be found in the AEA RCT Registry (AEARCTR-0002028). The reader can access the survey information sheet and interactively follow the survey at `http://ssd.az1.qualtrics.com/jfe/form/SV_8CVUQyyMcdITPHD`.

[25]Subjects who instead supported the law (N = 152) did not participate in the third part of the survey.

alternative interpretation of this result would be that participants might judge the Swiss person less negatively when a majority of Swiss people support the ban, *regardless of whether his support was expressed in public.* For example, it could be that participants feel that they cannot blame a person for privately holding a view if that person is surrounded by many other people who also hold that view and who could have influenced this person's convictions. With similar implications, participants might change their own opinion about minarets after learning that a majority of Swiss people are against them, and for this reason start judging the Swiss participant less negatively for privately holding these same views.

To explicitly rule out these possibilities, in the days immediately following wave 1, we conducted an experiment with a slightly modified version of the protocol. In this second wave, participants were informed about the fact that the 24-year-old male from Switzerland expressed his opinion in an *anonymous* survey. To make sure we could hire enough respondents, in this wave we recruited participants from the twelve states in which Hillary Clinton won the presidential election with the highest margin (California, Hawaii, Maryland, Massachusetts, New York, and Vermont as in wave 1, plus Connecticut, Delaware, Illinois, New Jersey, Rhode Island, and Washington).[26]

The design of this experiment was almost identical to the original version. Once again, we focus on subjects who reported to be against the introduction of the ban (N = 427).[27] The main difference with the original version is that we emphasized that the Swiss participant expressed his opinion anonymously. Both in the control and in the treatment conditions, instead of writing, as before, that "we matched you with a participant from another survey," in this version we wrote "we matched you with a participant from another anonymous survey." In our treatment groups we emphasized once again that the survey the Swiss person participated in was anonymous: "In our anonymous survey, like the one you just completed, he said he supports the prohibition of the building of minarets in Switzerland." We call this first treatment group the *anonymous anti-minarets* group. Finally, instead of writing "like 57.5% of Swiss respondents, the participant supports the prohibition of the building of minarets in Switzerland," in this case we wrote "According to numbers from 2009, 57.5% of Swiss respondents are in favor of prohibiting the building of minarets." We call this second treatment group the *anonymous anti-minarets, public support* group.[28]

---

[26]As in the other experiments, MTurk workers with at least 80% approval rate could see our request, which in this case was described as a "4-5 minutes short survey" with a reward of $0.50. Each worker could participate in the survey only once, and only if he/she did not participate in our other experiment. Workers who clicked on the request were displayed detailed instructions about the task, and given access to links to the study information sheet and the actual survey. The survey was conducted on the online platform *Qualtrics*. The reader can access the survey information sheet and interactively follow the survey at `http://ssd.az1.qualtrics.com/jfe/form/SV_1S5OypQLiIhlsPj`.

[27]Subjects who instead supported the law (N = 138) did not participate in the third part of the survey.

[28]Our design also included a fourth group (N=136 in wave 1, and N=139 in wave 2), where participants were instead told: "Building minarets is illegal in Switzerland, following a 2009 referendum. Like 57.5% of Swiss respondents, the participant supports the prohibition of the building of minarets in Switzerland. However, he did not vote in the referendum since he was under legal voting age" in wave 1, and "In our anonymous survey, like the one you just completed, he said he supports the prohibition of the building of minarets in Switzerland. Building minarets is illegal in Switzerland, following a 2009 referendum. According to numbers from 2009, 57.5% of Swiss respondents

E.1.1 Beliefs about the Swiss Ban

At the end of the intervention, subjects in the control group were also asked about their beliefs regarding the share of the Swiss who supported banning the construction of minarets, and whether they believed the ban is legal in Switzerland. In the first wave we did not collect this information for individuals in the *anti-minarets* and *anti-minarets public support* groups. To check whether their beliefs about the share of the Swiss population supporting the ban are changed by the treatments, we included these questions for both the control group and the treatment groups in the second wave. The share of those thinking that a majority of the Swiss support the ban is almost identical in the control group and the *anti-minarets* group (respectively 20 and 25%, with a *p*-value for the test of equality of 0.301), but increases to 63% in the *anonymous anti-minarets public support* group (*p*-values of the test of equality are less than 0.001 for either groups). The median belief about the share of the Swiss population supporting the ban is 30% in both control and *anonymous anti-minarets* groups, and 55% in the *anonymous anti-minarets public support* group. This confirms that our experimental manipulation indeed shifted beliefs about the level of popular support for the ban in Switzerland.[29]

Participants across conditions were also asked whether they believed the construction of minarets is legal in Switzerland: in all three groups, a majority reported to think that constructing minarets was legal (88% in the control group, 77% in the *anti-minarets* group, and 74% in the *anti-minarets public support* group).[30]We can thus rule out that the effects are affected by the fact that the ban is enacted as law, and can thus isolate the role of pandering to the public opinion on participants' judgment of the Swiss player.

E.2 Results

Figure D1 displays the main findings from Experiment 2B. Panel A displays comparisons of average donations across groups. In the control condition, where participants were only told that they are matched with a 24-year-old male from Switzerland, we observe an average transfer to the

---

are in favor of prohibiting the building of minarets. However, the person you are matched with did not vote in the referendum since he was under legal voting age" in wave 2. This *anti-minarets, referendum* treatment was intended to test whether providing information a view that is not only held by a majority but is also *official* would further change the donation rates. We found no effect of this additional treatment relative to the second treatment group, neither in the original version nor in the anonymous version of Experiment 2B, suggesting that institutionalization/legitimacy also does not seem to play a role on the receiver's side.

[29]Here we report the numbers from the second wave of the experiment, since the first wave only asked beliefs for the control group. The numbers for this group are very similar across waves. In the first wave, 17% of control group participants believe a majority of Swiss people support the ban, compared to 20% in the second wave. The median belief is 30% for the control groups in both waves.

[30]While the beliefs are significantly different when comparing the control group with either of the two treatment groups (the *p*-values for the test of equality are 0.013 against the *anonymous anti-minarets* group and 0.002 against the *anonymous anti-minarets public support group*), there is no statistical difference between the two treatment groups (the *p*-value for the test of equality is 0.500).

Swiss participant of $1.03. The average transfer is substantially lower for subjects in the *anti-minarets* group, who are also told that this person supports the prohibition on building minarets in Switzerland: the average transfer for this group is $0.69. The effect of informing subject about the anti-Muslim views of the Swiss participant is statistically significant (p-value<0.001). However, the average transfer among subjects in the *anti-minarets, public support* group who are told that the majority of Swiss respondents are against minarets is $0.92, which is not statistically different from the average transfer in the control group (the p-value of the difference is 0.162) but is substantially higher than the average transfer in the *anti-minarets* group (p-value=0.013). The average donation in the *anonymous anti-minarets* group is identical to that in the *anti-minarets* group, at $0.69. The average donation in the *anonymous anti-minarets, public support* group is also very similar: $0.70. These two levels are significantly different from the average in the control group (p-value<0.001 in both cases). The average donation in the *anonymous anti-minarets, public support* group is also significantly lower than the one in the first wave version of the treatment (p-value=0.014).

Panel B compares the share of participants who do *not* share anything from their $3 endowment with the Swiss person. The percentage of participants deciding not to transfer anything to the Swiss respondent increases from 22% in the control group to 42% in the *anti-minarets* group (p-value<0.001), while only 27% of subjects in the *anti-minarets, public support* decide to keep all $3. This percentage is not statistically different from the one in the control group (p-value=0.370), but is substantially lower than the one for subjects in the *anti-minarets* group (p-value=0.013). Here again, the levels of the outcome variable in the two anonymous treatments are almost identical to the level in the *anti-minarets* group: 43% and 44%. Importantly, the share of participants not donating is significantly higher in the *anonymous anti-minarets, public support* group when compared to the non-anonymous version of the treatment (p-value=0.004).

Figure D1: **Experiment 2B: Donation Rates**

Panel A   Panel B
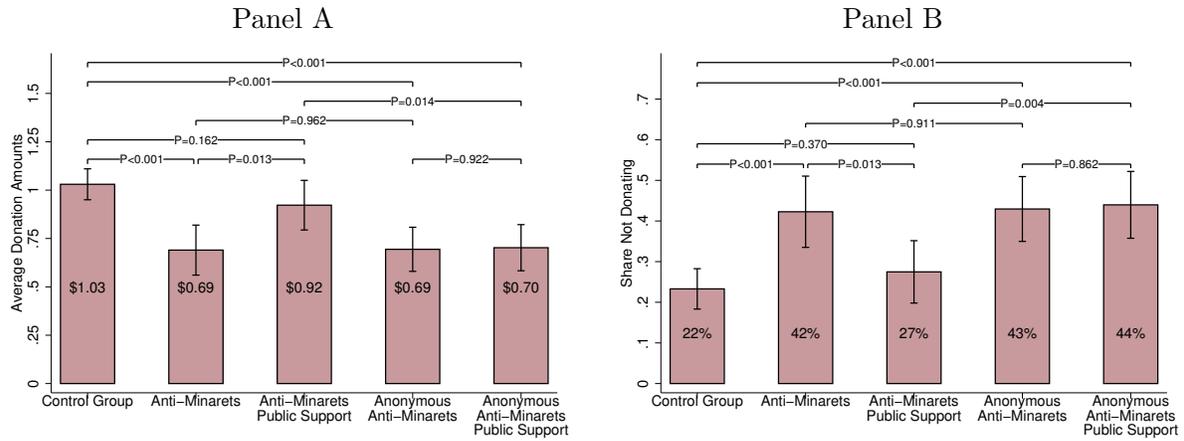
Notes: Panel A displays average donation amounts to the Swiss individual in the five experimental conditions: the control group (N = 279, pooling 142 observations from the first version of Experiment 2 and 137 observations from the second anonymous version of Experiment 2), the *anti-minarets* group (N=133), and the *anti-minarets public support* group (N=131), the *anonymous anti-minarets* group (N=149), and the *anonymous anti-minarets public support* group (N=141). Panel B displays the percent of subjects not making positive donations. Error bars reflect 95% confidence intervals. Top horizontal bars show *p*-values for *t* tests of equality of means between different experimental conditions.

F   Experiment 3: Expressing Islamophobia

F.1   Experimental Design

The design of Experiment 3 is very similar to Experiment 1: it uses donation decisions made either in a private or in a public condition to study the social acceptability of a view. The main difference with respect to Experiment 1 is that instead of varying the perceived local popularity of candidate Trump, we directly randomize the perceived local popularity of anti-Muslim sentiments. This allows us to test whether the effects documented in Experiment 1 are related to updates in the popularity of anti-immigration sentiments, or if instead the updates in the popularity of Donald Trump affect donation decisions through some other mechanism. Moreover, we also focus on the role of the legitimacy of a view in determining its social acceptability. One might also be concerned that, in Experiment 1B, participants might expect xenophobic policies to be institutionalized under Donald Trump's administration (and believe that such expectation is also shared by the surveyors). Such institutionalization/legitimacy could potentially increase the social acceptability of xenophobia. For this purpose, we include a treatment in which we inform subjects about the fact that a certain policy is unconstitutional.

Given our previous findings that the wedge between private and public donations to the *Federation for American Immigration Reform* had disappeared after the presidential election in the six originally studied states (and our overall concern that the social acceptability of xenophobia had increased in the country as a whole), we made three additional changes to the protocol in Experiment 1B: we expanded the set of states in our recruitment of participants, referred to stronger xenophobic (here, Islamophobic) language, and included an organization with relatively more extreme views.[31]

Specifically, in early February 2017, we recruited participants (N = 574) from all the states in which Donald Trump won the presidential election. MTurk workers with at least 80% approval rate could see our request, which was described as a "5 minute survey" with a reward of $0.50. Each worker could participate in the survey only once. Workers who clicked on the request were displayed detailed instructions about the task, and given access to links to the study information sheet and the actual survey. The survey was conducted on the online platform *Qualtrics*.

After answering a number of demographic questions, a third of the participants were randomly informed about the fact that a large share of respondents of an anonymous online survey supported the ban of Muslims from public office (*public support information* condition):[32]

"In a recent anonymous survey we conducted online, we found that a **very large**

---

[31]The experiment can be found in the AEA RCT Registry (AEARCTR-0001994). The reader can access the survey information sheet and interactively follow the survey at `https://ssd.az1.qualtrics.com/jfe/form/SV_beWUpftLZi3zW5f`.

[32]We used information from a previous anonymous survey we conducted on MTurk (N = 96) in which 42% of the respondents expressed support for that ban: to participate MTurk workers had to have an approval rate of at least 80% and to identify themselves as conservatives.

proportion of respondents think that Muslims should be prohibited from holding public office. This suggests that there is popular support for this type of ban."[33]

Another third were additionally informed about the fact that such a ban is unconstitutional and that Donald Trump would not be able to enact it (*unconstitutionality information* condition):

"Regardless of popular support, prohibiting Muslims from holding public office is **unconstitutional** and will not be enacted. The 5th and 14th Amendments imply that state and federal governments cannot discriminate against employees or job applicants on religious grounds. This means that President Donald Trump will not enact this type of ban."

The remaining third were not given any information (*control* condition).

Participants were then asked to predict the share of individuals who would they think would say in an anonymous online survey that they think Muslims should be prohibited from holding public office. This provides a measure of the perceived popularity of anti-Muslim policies.

In the next part of the intervention, we measured the perceived social acceptability of expressing strong anti-Muslim sentiment using a donation experiment with real stakes. Participants were first told that they would be given the opportunity to make a donation to a randomly drawn organization that could either be anti-Muslim or pro-immigration, to ensure that participants would not associate the experimenters with a specific political view. To maximize power and avoid direct deception, the randomization was such that more than 99% of participants (N=573) would get assigned the organization we were interested in: *ACT for America*.[34] To make sure that the participants were aware of the organization's very strong anti-immigration stance, a few more details about the organization and its founder were provided in the experiment:

ACT for America is the largest grassroots **anti-Muslim** organization in the U.S actively working to promote anti-Muslim legislation and opinion. The founder of ACT for America is Brigitte Gabriel, the author of a book titled 'They Must Be Stopped' and who argued that **Muslims should be prohibited from holding public office** because "a practicing Muslim, who believes in the teachings of the Koran, cannot be a loyal citizen of the United States." ACT for America believes that Muslims represent a threat to both national security and American values; its Thin Blue Line project comprehensively mapped the addresses of U.S. Muslim student associations and other Islamic institutions as sites of national security concern.

[33]To avoid deception, we used the vague expression "very large proportion," which does not imply that a majority of respondents held that position.

[34]The pro-immigration organization was once again the *National Immigration Forum*.

Participants were then asked if they would like to authorize the researchers to donate $1 to that organization on their behalf. The money would not come from the subject's $0.50 payment for participation in the study. Moreover, the participant would also be paid an *extra* $1 (or about 1/6 of an hourly wage on MTurk) if he/she authorized the donation. Rejecting the donation would not affect the monetary payoffs to the participants in any way other than through the loss of this extra amount.

In addition to the original randomization of informing subjects about the popularity and unconstitutionality of the ban, we introduced a second layer of cross-randomization at the donation stage. Half of the participants were assured that their donation authorization would be kept completely anonymous, and that no one, not even the researchers would be able to match their decision to their name: we refer to this condition as the *private* condition. The other half of the subjects were instead informed, right before the donation question was displayed to them, that they might be personally contacted by the research team to verify their answers to the questions in the remaining part of the survey: this is what we refer to as the *public* condition.

F.2 Results

Appendix Figure E1 displays the main findings from Experiment 3. In the control condition, we observe, like in Experiment 1, a wedge between donation rates in private and in public: a drop from 40% in private to 29% in public (the p-value of a t test of equality is 0.099). Among individuals in the public support information condition, we observe no difference in private and public donation rates, which are 35% and 37%, respectively (p-value=0.757). These results are very similar to the results in Experiment 1 (although we use different population, organization, and treatment), and indicate that the information provided on the popularity of anti-Muslim sentiments causally increased the social acceptability of the action to the point of eliminating the original social stigma associated with it. Among individuals in the unconstitutionality information condition, we again observe no difference in private donation rates, which are 29% and 30% respectively (p-value=0.795).

However, we find a difference in private donation rates between the unconstitutionality information and control conditions (p-value=0.083), suggesting that the information is possibly decreasing privately-held support for the Islamophobic policy.

Both information conditions positively update average beliefs about the popularity of the anti-Muslim policy when compared to the control group. In the control group, the average guess was that 45% of respondents of an online anonymous survey would support the anti-Muslim policy. The average went up to 48% in the unconstitutionality information condition (p-value=0.183 against the control group) and to 52% in popular support information condition (p-value=0.004 when compared to the control group). This is consistent with subjects informed about the unconstitutionality of banning Muslims from public office also reducing their beliefs about the popularity of the policy.

Taken together, these results suggest that the positive update in the perceived popularity of

the Islamophobic policy reduces the wedge in private vs public donations and confirms that the channel of legality/institutionalization or other channels related to the popularity of Donald Trump but not to the popularity of xenophobic sentiments are not the main drivers of our findings.

Figure E1: **Experiment 3: Donation Rates**
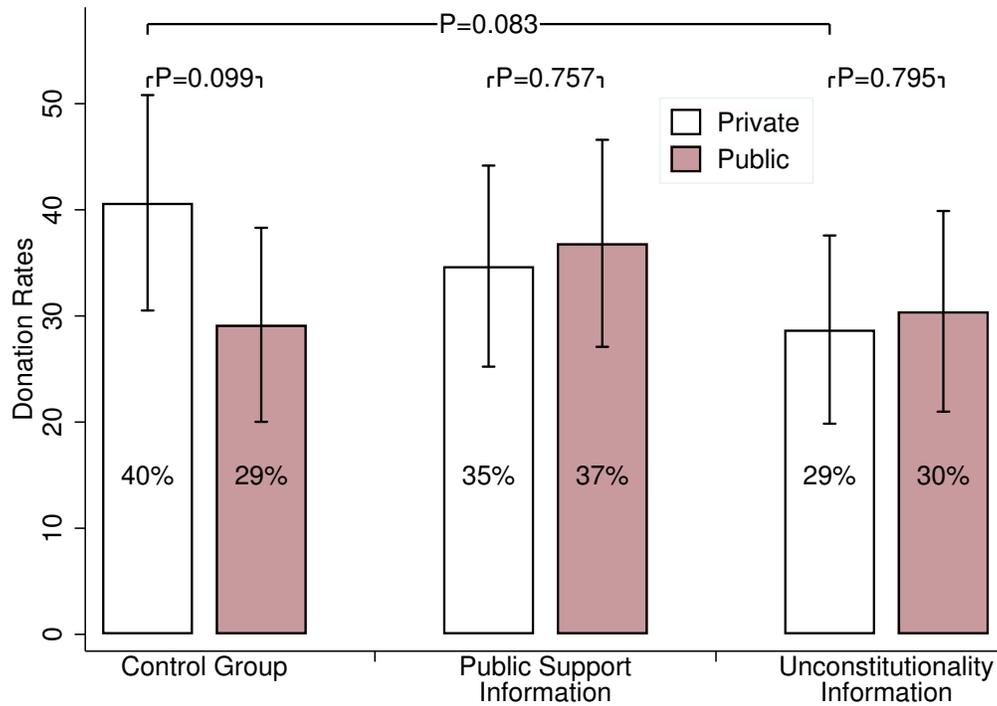
Notes: the two bars on the left display donation rates to the anti-Muslim organization for individuals in the private and public conditions in the *control* group (respectively N=91 and N=96), the two central bars display those in the *public support information* group (respectively N=98 and N=95), and the last two bars display those in the *unconstitutionality information* group (respectively N=101 and N=92). Error bars reflect 95% confidence intervals. Top horizontal bars show *p*-values for *t* tests of equality of means between different experimental conditions.