DISGUISING PREJUDICE:
POPULAR RATIONALES AS EXCUSES FOR INTOLERANT EXPRESSION

Leonardo Bursztyn
Ingar K. Haaland
Aakaash Rao
Christopher P. Roth

ABSTRACT

We study the use of popular rationales to justify public anti-minority actions. Rationales to oppose minorities change some people's private opinions, leading them to take anti-minority actions even if they are not prejudiced against minorities. When these rationales become widespread, prejudiced people can pool with unprejudiced people who are persuaded, decreasing the stigma associated with anti-minority expression and enabling greater public opposition to minority groups. We examine this mechanism through several large-scale experiments in the context of anti-immigrant behavior in the United States. In a first experiment, subjects learn that a previous respondent authorized a donation to an anti-immigrant organization and then make an inference about the respondent's underlying motivations. Subjects informed that their matched respondent learned about a study claiming that immigrants increase crime rates before authorizing the donation see the respondent as less intolerant and more easily persuadable. In a second experiment, subjects learn about that same study and then choose whether to authorize a publicly observable donation to the anti-immigrant organization. Subjects who are informed that their exposure to the rationale will be publicly observable are substantially more likely to make the donation than subjects who are informed that their exposure will remain private. Our findings suggest that prominent public figures can lower the social cost of intolerant expression by popularizing rationales, contributing to waves of anti-minority behavior.

Leonardo Bursztyn
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER
bursztyn@uchicago.edu

Ingar K. Haaland
Department of Economics
University of Bergen
Fosswinckels gate 14
Bergen, Norway
ingar.haaland@uib.no

Aakaash Rao
Harvard University
Department of Economics
Littauer Center of Public Administration
1805 Cambridge St
Cambridge, MA 02138
arao@g.harvard.edu

Christopher P. Roth
Department of Economics
University of Warwick
Coventry CV4 7AL
United Kingdom
Christopher.Roth@warwick.ac.uk

1 Introduction

As leaders of many democracies and popular segments of the media embrace a nativist, anti-immigrant rhetoric, anti-minority expression has been on the rise (United Nations, 2020). A growing body of historical evidence suggests that such rhetoric can have substantial effects on people's willingness to engage in xenophobic actions against immigrants and other minorities (Enikolopov and Petrova, 2015; Zhuravskaya et al., forthcoming). For example, radio propaganda contributed to increased killings during the Rwandan genocide (Yanagizawa-Drott, 2014) and anti-Semitic expression in Nazi Germany (Adena et al., 2015).[1]

While these studies identify a causal effect of mass messaging on behavior, the underlying mechanisms remain unclear. The most-studied channel is persuasion: one might attribute the growing wave of anti-immigrant rhetoric and violence in the United States to private attitudes toward immigrants growing more negative.[2] Yet survey evidence suggests quite the opposite: both Democrats and Republicans reported feeling, if anything, more *warmly* toward both legal and illegal immigrants in 2018 than in 2014 (Gonzalez-Barrera and Connor, 2019). Consistent with this observation, recent experimental work finds relatively small or null effects of information on immigration policy preferences (Hopkins et al., 2019; Alesina et al., 2019; Grigorieff et al., 2020). Together, the survey and experimental evidence suggest that mechanisms beyond persuasion may be driving trends in public anti-minority expression.

In this paper, we propose an additional mechanism through which the spread of anti-minority narratives might affect public behavior. By popularizing rationales to oppose minorities, whether true or false, public figures can enable intolerant expression by lowering the social cost of publicly expressing otherwise-stigmatized positions. For example, consider people who oppose immigration from Mexico simply because they dislike Mexicans, yet cannot express this opposition in a public setting without incurring social costs. Once an anti-Mexican rationale becomes common knowledge (e.g., a politician claims that Mexican immigrants bring crime into the country), these people are given an excuse: they can attribute their position to a belief that Mexican immigrants commit more crime, even if they privately do not believe the rationale is true.[3] The key point is that the

---

[1] For a review of social and economic impacts of media, see DellaVigna and La Ferrara (2015).

[2] For example, the number of white nationalist hate groups in the United States has grown by 55 percent since January 2017. ("White Nationalist Hate Groups Have Grown 55% in Trump Era, Report Finds." *The Guardian*, March 18, 2020.) Islamophobic rhetoric among elected officials at all levels of government has also increased substantially. ("Islamophobia in the US: It Goes Way beyond Trump." *Vox News*, April 6, 2018.)

[3] In many situations, this mechanism amounts to disguising taste-based discrimination (Becker, 1957) as statistical discrimination (Phelps, 1972; Arrow, 1973).

availability of the excuse opens up explanations — other than intolerance — for their anti-Mexican positions, reducing the extent to which they expect observers to update about their intolerance and thus increasing their willingness to express their private views. Thus, even if the rationale has no direct persuasive impact, it can serve as an excuse as long as it is plausible that *others* might be persuaded.

In the United States, the Trump administration has used a variety of rationales to justify the construction of the US–Mexico border wall: among them, the claim that Mexican immigrants are violent criminals and rapists, that terror suspects are pouring into the US over the Mexico border, and that undocumented immigrants take jobs from American workers.[4] These rationales allowed the administration to maintain plausible deniability about the motives underlying controversial policies: when asked about Trump's immigration policies, 49 percent of voters answered that they were motivated by a "sincere interest in controlling our borders", while 41 percent answered that they are motivated by "racist beliefs" (Snow, 2019). More recently, the Trump administration used the COVID-19 pandemic to provide widespread justifications for xenophobic expression against immigrants and asylum seekers.[5] The results of a survey from May 2020 suggest a potentially important role of this rationale in enabling anti-immigrant expression: Republicans were 19 percentage points more likely to publicly support a *permanent* ban on Mexican immigration when they were able to publicly attribute their opposition to Mexican immigration to the threat of immigrants bringing COVID-19 into the country (compared to a base rate of 32 percent).[6] Other populist leaders have also used similar rationales to target minority groups: Marine Le Pen, Boris Johnson, and Narendra Modi have all popularized (unsubstantiated) claims that migrants and refugees disproportionately commit crimes and take jobs from native workers in order to justify stricter immigration laws and crackdowns on immigrants already in the country, while Matteo Salvini has widely spread the unsupported claim that Italy faces an "invasion" of African mafia bosses in order to justify anti-refugee policies.[7]

Uncorroborated rationales have also been widely used to justify discrimination against other

---

[4]See "A Quick History of Trump's Evolving Justifications for the Border Wall." Morin, Rebecca, *Politico*, January 8, 2019.

[5]See, for example, "The Trump administration is using the pandemic as an excuse to target immigrants and asylum seekers." *Vox News*, May 15, 2020.

[6]For further details, see Appendix C.

[7]See, for example: "In Their Own Words: Marine Le Pen and Emmanuel Macron." Joseph, Yonette, *The New York Times*, May 5, 2017. "Boris Johnson Follows Donald Trump's Lead On Immigration, Unveiling Plans For Major Crackdown." Da Silva, Chantal, *Newsweek*, November 18, 2019. "Modi Defends Indian Citizenship Law Amid Violent Protests." Schultz, Kai, *The New York Times*, December 22, 2019. " 'Migrant menace': Salvini accused of targeting refugees and ignoring mafia." Tondo, Lorenzo, *The Guardian*, February 12, 2019.

minority groups. Since Reconstruction and continuing to this day, political parties in the United States have exaggerated the severity of voter fraud at the ballot box — a "vanishingly rare" phenomenon (Wines and Blinder, 2016) — to justify measures that disproportionately depress minority turnout (Cohen, 2012). US politicians and military leaders have for decades cited concerns about "unit cohesion" as an excuse to justify discriminatory policies targeting minorities in the military: African-Americans, women, gays and lesbians, and most recently transgender individuals (Frank, 2009).[8] The Reagan campaign popularized the image of the "welfare queen" to appeal to racist stereotypes about African-American single mothers (Mendelberg, 2001).

To illustrate the intuition behind the excuse effect and motivate our experiments, consider a setting with two agents: a sender and a receiver.[9] Both agents are first exposed to an anti-immigrant rationale. The sender then chooses whether to engage in an anti-immigration action (for example, donating to an anti-immigrant organization). The receiver observes the sender's decision and makes an inference about the sender's motives. Agents differ on two privately-known dimensions. First, some of them are tolerant toward immigrants, while others are not. Second, some agents are persuaded to donate when they learn the rationale — *regardless of their tolerance type*.[10] Senders receive expressive utility from making a donation decision consistent with their own tolerance type as well as social utility from leading the receiver to believe that they are of the same tolerance type.

Consider a situation in which the rationale for donating to the anti-immigrant organization is privately known to both the sender and the receiver, but the receiver is unaware that the sender knows the rationale. The receiver will infer with certainty that the sender is intolerant if he decides to donate, and the sender, knowing this, may therefore strategically choose not to donate, even if he is privately intolerant. This inference changes if the receiver believes that the sender has been exposed to the rationale (e.g. if the rationale is widely known in the population) since the receiver understands there are two potential reasons the sender might have donated. In particular, the sender might be intolerant, or the sender might be *persuadable* (e.g., the receiver might think the sender is gullible or that the sender is particularly worried about the consequences associated with

---

[8]See, for example, "Unit cohesion and how it's been used to justify discrimination in the military." *ABC News*, July 27, 2017. Frank (2009) documents how military leaders popularized the "unit cohesion" rationale to disguise the true motivation for the controversial "don't ask, don't tell policy" used to prevent gays and lesbians from openly serving in the army: namely, moral opposition to "deviant sexual behavior".

[9]We formalize this intuition with a simple model in Section 2.

[10]Differences in persuasion may arise from differences in the extent to which information shapes beliefs (for example some agents may be more gullible than others) *or* because some agents care more about the consequences associated with the rationale (for example, low-skilled natives who worry more than high-skilled natives that increased low-skilled immigration will threaten their employment prospects).

the rationale). To the extent that being persuadable is less stigmatized then being intolerant, the popular knowledge of the justification reduces the social cost of donating and increases donation rates by allowing intolerant senders to pool with agents with a "good reason to donate" – in essence, pretending that their motivation for donating stems from the anti-immigrant rationale rather than their intolerance.[11] Motivated by this framework, we present two main experiments examining the use of rationales to justify anti-immigrant positions.

In Experiment 1, conducted with a broadly representative sample of 3,047 Democrats, we study whether the availability of an excuse influences how respondents interpret xenophobic actions, focusing on one of the most widely-used excuses for anti-immigrant expression: the claim that immigrants disproportionately commit violent crime.[12] In particular, we match participants with a respondent from previous study who authorized a donation to "Fund the Wall," an organization working to fund the proposed wall along the US–Mexico border. We begin by informing participants about a recent study (Lott, 2018), which claims that undocumented immigrants in Arizona commit crimes at substantially higher rates than comparable US citizens.[13] Our key experimental variation is to vary whether our subjects believe that their matched respondent knew about the study before making their decision: subjects in the *Excuse* condition are matched with a respondent who knew about the study before making their decision, while subjects in the *No Excuse* condition are matched with a respondent who did not know about the study. We investigate whether subjects infer that participants who had this "excuse" for donating are less intolerant than participants who donated without an excuse. We also examine whether subjects infer that participants who had an excuse for

---

[11]Justifications which are widely known thus work best as excuses. It is certainly true that people who express intolerant views can, when challenged, claim to believe a fringe theory that justifies these views, the existence of which the majority of the population is unaware. But this is true only if they have the opportunity to explain themselves to their audience: in the absence of a justification, their audience will believe that their xenophobic expression is due to innate xenophobia, not to persuadability. Thus, fringe narratives are limited in the extent to which they can be effectively used as excuses.

[12]We are particularly interested in how excuses affect judgment vis-a-vis an audience that *disapproves* of the action, as this is precisely the audience before which an agent may require an excuse. We thus focus on Democrats, who are most likely to disapprove of the decision to donate to Fund the Wall.

[13]The Trump administration has cited this study repeatedly as evidence for the impact of illegal immigration on crime. For example, in a January 2018 speech on "national security and immigration priorities of the administration," then-Attorney General Jeff Sessions claimed that the study proved that "tens of thousands of crimes have been committed in this country that would never have happened if our immigration laws were enforced and respected like they ought to be". (Sessions, Jeff. "Attorney General Sessions Delivers Remarks on National Security and Immigration Priorities of the Administration." *Justice News*, January 26, 2018.) We also inform respondents that many researchers have challenged the study's validity (Nowrasteh, 2018), and to further ensure that they are not left with a distorted view of the relationship between immigration and crime, we provide respondents with a short summary of the empirical evidence on the effects of immigration on crime and a link to a relevant meta-analysis at the end of the experiment (Ousey and Kubrin, 2018), which highlights, if anything, a negative association between increases in immigration and crime.

donating did so for another reason — because they are gullible and were persuaded by the study despite its methodological flaws.

To most closely capture the natural process of inference and to avoid priming respondents, we first measure participants' beliefs about their matched respondents' motives for donating to Fund the Wall using an open-ended question, directly measuring what "comes to mind" through a pre-registered text analysis procedure. We then turn to more structured measures of beliefs: half of the participants make an incentivized guess about their matched respondent's score on a test measuring cultural tolerance, while the other half make an incentivized guess about their matched respondent's score on a test measuring gullibility.[14]

We find strong treatment effects on both measures of type inference. In describing why they believed their matched respondent chose to donate to Fund the Wall, participants matched with a respondent who had no excuse for donating are 7 percentage points (70%) more likely to use a word related to intolerance ($p < 0.001$) and 3 percentage points (43%) less likely to use a word related to gullibility ($p < 0.001$) relative to participants matched with a respondent who had an excuse. We find similar treatment effects on the structured belief measures: participants believe that a matched respondent with an excuse scored 0.14 standard deviations lower on the intolerance scale ($p < 0.001$) and 0.32 standard deviations higher score on the gullibility scale ($p < 0.001$). Taken together, our evidence suggests that publicly known rationales for xenophobic behavior strongly influence how an audience updates about the underlying motives. In an auxiliary experiment, we show that respondents are less likely to punish gullible partners than intolerant partners, consistent with our assumption that xenophobia is indeed more socially sanctioned than persuadability.

In Experiment 2, we investigate whether people exploit widespread rationales to avoid the social stigma associated with publicly expressing intolerant views. Specifically, we recruit a broadly representative sample of 3,728 Republicans and Independents and examine whether they are more willing to publicly undertake an anti-immigrant action — authorizing a donation to Fund the Wall — when they have an excuse available. We inform all participants about the Lott (2018) study, which claims that undocumented immigrants commit crimes at much higher rates than US citizens, and then give participants the opportunity to authorize a $1 donation to Fund the Wall. We tell

---

[14]Of course, gullibility is only one of a set of possible "second types." Holding fixed the extent to which two agents believe the rationale and those agents' tolerance, one agent may choose to donate while the other does not because she is more risk-averse, because she will be more affected in the event that the rationale is true, or because her costs of donating are lower, among other potential reasons. We focus on this single "second type" to discipline our exercise for a number of reasons: it is (arguably) the most natural "second type," it was the most frequent reason cited in our pilot results, and it is more objectively coded than other possible types.

participants that we will post their individual donation decisions on our website when the study is published in an academic journal, and that in order to communicate our research findings to the public, we will publicize the website among residents in their city. This generates a real social cost of authorizing a donation, particularly in areas where respondents expect the majority of the population to disapprove of the donation.

Identifying the "excuse effect" requires disentangling it both from the direct effect of persuasion ("first-order" persuasion) and from a change in anticipated social approval associated with changes in the audience's beliefs ("second-order" persuasion). We hold first- and second-order persuasion fixed across the *Excuse* and *No Excuse* condition; thus, the key treatment varies only the availability of an excuse for donating. In particular, participants in the *Excuse* condition see that their audience will learn that they knew about the Lott study when making the donation decision, while participants in the *No Excuse* condition believe that their audience will *not* know that they knew about the Lott study.

We find a large and statistically significant excuse effect on participants' willingness to publicly donate to Fund the Wall. Participants in the *Excuse* condition are 6.3 percentage points (13%) more likely to authorize the donation than respondents in the *No Excuse* condition ($p < 0.001$). To benchmark the effect size, we compare the donation rate in a control condition — in which participants are not informed about the study — with the donation rate in the *No Excuse* condition, which allows us to identify the joint effect of first- and second-order persuasion. We find that this joint effect is small relative to the "excuse effect." This again suggests a quantitatively important role of commonly known excuses relative to the direct and indirect effects of persuasion. Moreover, the effect is driven by participants who live in more liberal areas, suggesting that participants more strongly require excuses when their audience is likely to disapprove of their actions. Evidence from a number of different exercises, and a successful replication of our findings with a more subtle treatment manipulation using a sample of 1,373 respondents in October 2020, suggests that experimenter demand effects are not driving our results.

Related Literature Our paper builds on theoretical literature on the effects of social image concerns on economic and moral decision-making (e.g. Bénabou and Tirole 2006). Related to our work is Bénabou et al. (2018), which presents a theoretical model of the production and circulation of arguments justifying actions on the basis of morality. We also build on a growing empirical literature studying the effect of social image concerns on political and economic outcomes (including

moral behavior, as in Ariely et al. 2009; Lacetera and Macis 2010; Ewers and Zimmermann 2015; voting, as in DellaVigna et al. 2017; tax evasion, as in Perez-Truglia and Troiano 2018; Besley et al. 2019; identity choice, as in Jia and Persson 2019; campaign donations, as in Perez-Truglia and Cruces 2017; educational investments, as in Bursztyn and Jensen 2015; and labor market choices, as in Bursztyn et al. 2017). Relative to existing work, which generally highlights a single type dimension on which respondents signal and update (Bénabou and Tirole, 2006), a key contribution of this paper is to show that people can strategically use information to influence how others will assess their motives on *two* dimensions, with important consequences for publicly-observable behavior. Thus, in contrast to previous work showing that one's beliefs about others' opinions matter for public behavior, we show that one's beliefs about how others will *update* about one's own motives also have significant effects on one's willingness to express an otherwise-stigmatized view. We therefore highlight the importance of commonly known rationales, which can be created by prominent public figures and the media. Our paper is thus related to laboratory evidence on strategic communication used to justify public actions (Foerster and van der Weele, 2018).

Our work also relates to a growing literature on social norms governing public behavior — in particular, to work examining how these norms unravel (Kuran, 1997; Bursztyn et al., 2020b,a). Our work is similar in that it examines how previously-stigmatized public behavior becomes socially acceptable, but it differs both conceptually and in its implications for equilibrium expression. Conceptually, we disentangle second-order beliefs (beliefs about others' views) from third-order beliefs (beliefs about others' beliefs about one's own motives) and show that the latter mechanism enables excuses to increase the public expression of intolerant positions by reducing the extent to which public expression is informative of private attitudes. Practically, the excuse mechanism allows even views that are *privately unpopular* — such as conspiracy theories about certain minorities — to be publicly expressed in equilibrium. In contrast, the social norms mechanism examined in Bursztyn et al. (2020a) and Bursztyn et al. (2020b) rests upon these stigmatized views already being widely held. Of course, the two mechanisms are mutually reinforcing. For example, intolerant views may initially emerge among only a small segment of the population, which may use excuses to lower the cost of publicly expressing these views to the rest of society. As a consequence of this public expression, others may then be privately persuaded. An event that serves to aggregate this private information — for example, an election, as studied in Bursztyn et al. (2020b) — can then bring previously-fringe views into the mainstream. Thus, excuses may thus be essential in facilitating the initial growth of stigmatized views by enabling them to be publicly expressed.

Several laboratory studies show that "moral wiggle room" can have substantial effects on behavior (Dana et al., 2007; Golman et al., 2017, 2016; Andreoni and Bernheim, 2009; Lazear et al., 2012; Hamman et al., 2010; Saccardo and Serra-Garcia, 2020; Exley, 2016; Cunningham and de Quidt, 2016). Because decisions in these settings are anonymous, these findings can be understood through a behavioral model of self-signaling, as in Bénabou and Tirole (2011a): people exploit moral wiggle room to take self-serving actions while convincing themselves that they are not acting selfishly. Our work differs from this literature in that we are interested in the role of excuses in justifying actions vis-a-vis *others* — we hold the self-excuse channel constant by exposing respondents to the same private information set — and we examine the implications of commonly-known rationales both for the interpretation of intolerant actions and for the decision to take these actions.[15]

The remainder of this paper proceeds as follows. Section 2 motivates our experiments with a simple theoretical framework. In Section 3, we present an online experiment examining how the availability of an excuse affects the interpretation of xenophobic actions. In Section 4, we present an online experiment showing that commonly known rationales increase xenophobic expression. We discuss policy implications and conclude in Section 5. The Appendix includes proofs of our theoretical results, additional tables and figures, a discussion of a motivating survey, and the full set of experimental instructions.

2   Conceptual Framework

To organize thoughts and motivate the design of our experiments, we present a simple model of communication that formalizes the strategic implications of a publicly known rationale for xenophobic expression. Our framework is closely related to the canonical model by Bénabou and Tirole (2006) on image concerns and other related models (Ali and Bénabou, 2020; Golman, 2020; Jia and Persson, 2019; Besley et al., 2019; Benabou and Tirole, 2011b). We abstract away from many of the considerations in these related models to highlight our key conceptual contribution: the implications of heterogeneity across *two* action-relevant type dimensions for equilibrium inference and behavior.

---

[15]A seminal contribution in psychology is Langer et al. (1978), which finds that subjects waiting to make Xerox copies were more likely to comply with a request when it was justified by a reason, irrespective of whether the reason was "bad" ("Excuse me...May I use the Xerox machine, because I have to make copies?") or "good" ("Excuse me...May I use the Xerox machine, because I'm in a rush). Langer et al. (1978) interprets this as evidence for the "mindlessness of ostensibly thoughtful action", arguing that people have simply been conditioned to comply with requests accompanied by justifications. Related work in psychology includes Bandura et al. (1996), Bandura et al. (2001), and Shalvi et al. (2015).

A society consists of a continuum of agents who differ on two dimensions. First, some are tolerant toward foreign cultures ($i = 0$), while others are intolerant ($i = 1$). Second, some are easily persuaded by the given rationale ("persuadables") whereas others are not. The two dimensions are independent; the probability that a given agent is tolerant is given by $p \in (0, 1)$, and the probability that a given agent is persuadable is $q \in (0, 1)$. Agents' individual types are private information, though the distribution of types is common knowledge. At the beginning of the game, two agents are randomly drawn from the society: one agent is the "sender" while the other is the "receiver." The sender and receiver are exposed to an anti-immigrant rationale. The sender can choose either to take an anti-immigrant action ($a = 1$) or not to take this action ($a = 0$).[16] The receiver observes the sender's action and makes an inference about whether the sender is tolerant or intolerant.

The persuadable sender is non-strategic, with actions characterized as follows: in the absence of viewing anti-immigrant information, the tolerant-persuadable sender does not take the anti-immigrant action, while the intolerant-persuadable sender takes the action. However, once exposed to anti-immigrant information, the tolerant-persuadable sender is persuaded and induced to take the action[17]; the intolerant-persuadable sender takes the action, as before.

The non-persuadable sender is strategic and receives social utility proportional to the receiver's belief that the receiver and sender share the same tolerance type. In particular, when the receiver believes with certainty that the sender is of the same tolerance type, the sender receives social utility $\bar{b}$, while when the receiver believes with certainty that the sender is of the opposite tolerance type, the sender receives social utility $\underline{b}$, with $\bar{b} > \underline{b}$. Given that the probability of being matched with a tolerant receiver is p and the probability of being matched with an intolerant receiver is $1 - p$, the sender's social utility from being perceived as tolerant with certainty is given by $b_0 := p\bar{b} + (1-p)\underline{b}$, while the sender's social utility from being perceived as intolerant with certainty is given by $b_1 := p\underline{b} + (1 - p)\bar{b}$. For simplicity, we assume that the sender's utility is not directly affected by the receiver's inference about the sender's persuadability. This assumption can be relaxed without affecting any of the main results below, so long as the image cost of being perceived

---

[16]Equivalently, we can interpret this as choosing between an anti-immigrant action ($a = 1$) and a pro-immigrant action ($a = 0$), or choosing between no action ($a = 1$) and a pro-immigrant action ($a = 0$). In our experimental setting, the anti-immigrant action corresponds to authorizing a \$1 donation to Fund the Wall, an organization that seeks to fund the proposed US-Mexico border wall.

[17]Differences in persuasion may arise because some people are more gullible than others, and thus the posterior probability that these gullible types assign to the event that the story is true shifts further from their prior than that of non-gullible types. Alternatively, these differences may arise because some people will be more affected if the state of the world implied by the rationale is true, and thus they are more willing to take the anti-immigrant action than other agents even if they assign the *same* probability to the event that the rationale is true as other agents. Said differently, differences in persuasion may arise from differences in belief updating or from differences in payoffs.

as persuadable is lower than the image cost of being perceived as intolerant.

Thus, the sender's expected social utility of inducing the receiver to believe with probability π that the sender is tolerant is given by $b(\pi) = \pi b_0 + (1 - \pi)b_1$. We assume that $p > 0.5$ such that $b_0 > b_1$, i.e. the expected social utility from being perceived as tolerant is strictly greater than the expected social utility from being perceived as intolerant.[18]

Both types of non-persuadable senders also receive expressive utility $v > 0$ from taking an action consistent with their tolerance type: in particular, the intolerant sender receives v when choosing to take the anti-immigrant action and 0 otherwise, while the tolerant sender receives v when they choose not to take the anti-immigrant action and 0 otherwise. The utility function of the non-persuadable sender with tolerance type $a = i$ is thus given as follows:

$$u_i\left(a, \pi\right) = v\mathbf{1}_{\{a=i\}} + \pi b_0 + (1 - \pi)b_1.$$

Let $\pi(a)$ denote the receiver's posterior belief that the sender is tolerant after observing the sender's action a. Then, the following proposition holds:

Proposition 1. *Non-persuadable senders' optimal actions are as follows*[19]*:*

$$a_0^*\left(\pi(\cdot)\right) = \mathbf{1}_{\left\{\pi(1)-\pi(0)>\frac{v}{b_0-b_1}\right\}}, \tag{1}$$

$$a_1^*\left(\pi(\cdot)\right) = \mathbf{1}_{\left\{\pi(1)-\pi(0)>-\frac{v}{b_0-b_1}\right\}}. \tag{2}$$

In other words, assuming $\pi(1) < \pi(0)$ (that is, the receiver believes senders who take the anti-immigrant action are more likely to be intolerant than senders who do not take the action), the tolerant non-persuadable agent never takes the action, while the intolerant non-persuadable agent takes the action if and only if the social image cost of doing so is small relative to the expressive utility. In what follows, we take the expressive utility v as given and endogenize π.

We consider the equilibria of two separate games, which map to our experimental conditions. In the *No Excuse (NE)* game, the receiver holds incorrect beliefs about the sender's information set (and this is known to the sender): the receiver believes with certainty that the sender did not see

---

[18]This assumption implies that the sender wants to be perceived as intolerant if they think their matched receiver is more likely to be intolerant than tolerant. Alternatively, we could assume that the sender always prefers to be perceived as tolerant irrespective of whether the receiver is more likely to be tolerant or intolerant. With $p > 0.5$, the model yields virtually identical results under this alternative assumption. That is, we can redefine $b_0 := \bar{b}$ and $b_1 := \underline{b}$ and the remainder of this section would be identical under this alternative assumption.

[19]We assume that the sender does not take the action when she is indifferent between taking and not taking the action; however, the results in the section do not depend on this assumption.

the anti-immigrant rationale prior to choosing her action. Thus, the receiver believes that there is no persuasion effect operating on the sender, and the receiver therefore believes with certainty that a sender who takes the action is intolerant, i.e. $\pi(a = 1) = 0$. In contrast, in the *Excuse* (E) game, the receiver (correctly) believes with certainty that the sender has seen the anti-immigrant rationale prior to choosing her action. Thus, the receiver no longer knows with certainty that a sender who takes the action is intolerant, since he knows he may be matched with a tolerant-persuadable sender who was persuaded by the anti-immigrant information to take the action.[20] Our solution concept for both games is Perfect Bayesian equilibrium in pure strategies, in which $\pi(\cdot)$ is consistent with each type of sender's actions and follows Bayes' rule when possible. We adopt the intuitive criterion to refine the set of off-path equilibria in the *Excuse* game (Cho and Kreps, 1987).[21]

The fact that the tolerant-nonpersuadable sender does not take the action in either game is immediate, since both social and expressive utility are strictly greater when the tolerant-nonpersuadable sender does not take the action than when she does.[22] When expressive utility v is small relative to social utility, the intolerant-nonpersuadable sender does not take the action either in the *Excuse* game or the *No Excuse* game because the social image costs of taking the action outweigh the expressive benefits. In contrast, when expressive utility v is large relative to social utility, the intolerant-nonpersuadable sender takes the action in both the *Excuse* game and the *No Excuse* game. For expressive utility v within a certain parameter range, there exists an equilibrium in which the intolerant-nonpersuadable sender does not take the action under the *No Excuse* game but takes the action under the *Excuse* game, assuming that the share of persuadable agents is suf-

---

[20]We view intolerance and gullibility as independent dimensions. An alternative interpretation might be that many agents are biased precisely *because* they are gullible: that is, they believe what they are exposed to, and they are simply exposed to more anti-minority information than pro-minority information. One implication is that providing the receiver with information relevant to the sender's type is likely to move the receiver's inference about the sender's bias and the sender's gullibility in the *same* direction. In Experiment 1, we find evidence that receivers' posteriors about senders' bias and gullibility, as measured by both an open-ended text analysis procedure and structured measures of beliefs, generally move in *opposite* directions, suggesting that this alternative channel may be limited in its quantitative importance. While the open-ended text may only capture the first motivation that comes to mind, the (incentivized) structured beliefs measures are not subject to this concern.

[21]In our model, persuadable and non-persuadable receivers are identical. In particular, tolerant-persuadable receivers who are persuaded by the anti-immigrant organization still judge intolerant senders in the same manner as tolerant-persuadable receivers, capturing the intuition that people care about the *motivations* behind others' actions. Moreover, persuadable receivers still use Bayes' rule to make inferences about the sender's motivations. We could alternatively model persuadable and non-persuadable receivers differently, such that persuadable receivers take senders' actions at face value: in other words, such that they believe with probability one that donors are intolerant and non-donors are tolerant. This alternative model would narrow the set of parameter values under which we observe our equilibria of interest, as described in Proposition 2, but would leave our model's predictions qualitatively unchanged.

[22]The fact that expressive utility from not donating is greater than from donating is by definition, while the fact that social utility from not donating is greater than social utility from donating follows from the assumption that $p > 0.5$.

ficiently large to allow intolerant-nonpersuadable agents to pool with tolerant-persuadable agents. We formalize this claim in Proposition 2, which we prove in Appendix A.

Proposition 2. *Suppose that*

$$\frac{(1-p)(b_0 - b_1)}{1 - qp} < v \leq \frac{p(b_0 - b_1)}{p + q(1-p)} \quad and \quad q < \frac{p^2}{2p^2 - 2p + 1}.$$

Then, there exists a unique equilibrium in the No Excuse game, and there exists a unique equilibrium in the Excuse game satisfying the intuitive criterion. The tolerant-nonpersuadable sender does not take the action in either game, while the intolerant-nonpersuadable sender takes the action only in the Excuse game.

Given the existence of the equilibrium as in Proposition 2, the following is an immediate corollary from the sender's equilibrium actions under the two conditions.

Corollary 1. *In the equilibria as in Proposition 2, the receiver's posterior belief that a sender who takes the action is intolerant is lower in the Excuse game than in the No Excuse game:*

$$1 - \pi^{NE}(1) = 1 > \frac{1-p}{1-pq} = 1 - \pi^{E}(1).$$

Moreover, the receiver's posterior belief that a sender who takes the action is non-persuadable is higher in the No Excuse game than in the Excuse condition:

$$\vartheta^{NE}(1) = 0 < \frac{q(1-p)}{1-qp} = \vartheta^{E}(1),$$

where $\vartheta(a)$ is the receiver's posterior belief after observing action a that the sender is non-persuadable.

The reasoning is straightforward: because the receiver believes that only the intolerant-persuadable sender takes the action in the *No Excuse* game, we have $\vartheta^{NE}(1) = 0$. In contrast, in the *Excuse* game, the receiver believes that intolerant-persuadable, tolerant-persuadable, and intolerant-sophisticated senders all take the action. Thus, we have $\vartheta^{E}(1) = \frac{q(1-p)}{(1-q)+q(1-p)} = \frac{q(1-p)}{1-qp}$.

3 Excuses and Interpreting Xenophobic Expression

We begin by examining how the availability of a rationale changes how an audience interprets the decision to donate to Fund the Wall, an organization that seeks to fund the proposed US–Mexico

border wall.[23] We are particularly interested in how excuses affect judgment vis-a-vis an audience that *disapproves* of the action, as this is precisely the audience before which an agent may require an excuse, and we thus focus on Democrats.[24]

In our framework, a "sender" may donate to Fund the Wall for two reasons. First, they may be intolerant. Alternatively, they may have been persuaded to donate after being exposed to the anti-immigrant rationale. The receiver observes the sender's donation decision and then uses this information to make an inference about the sender's motivations; the receiver may or may not be persuaded by the rationale.[25] Motivated by Corollary 1 of our framework, in Experiment 1, we study how the audience's inference about the sender's motivations is affected by the availability of an excuse.

3.1 Sample

We conducted Experiment 1 in partnership with the survey company Luc.id, a widely used online survey panel provider (Wood and Porter, 2019). We recruited a sample of 3,047 Democrats in February 2020.[26] Participants were directed to our survey on the online platform *Qualtrics*. Only participants who were over the age of 18, resided in the United States, indicated their consent to participate, and passed a simple test of attention were allowed to proceed. Our sample of respondents is broadly representative of Democrats in the United States (Appendix Table B1) and well-balanced on observables across treatment arms (Appendix Table B2). Experimental procedures and analyses were pre-registered in the AEA RCT Registry.

---

[23] All survey instruments are available in Appendix D.

[24] As of January 2019, 6 percent of Democrats or Democratic leaners favored "substantially expanding the wall", compared to 82 percent of Republicans or Republican leaners (Pew, 2019).

[25] Differences in persuasion may arise because some people are more gullible than others, and thus the posterior probability that these gullible types assign to the event that the story is true shifts further from their prior than that of non-gullible types. Alternatively, these differences may arise because some people will be more affected if the state of the world implied by the rationale is true, and thus they are more willing to donate than other agents even if they assign the *same* probability to the event that the rationale is true as other agents. Said differently, differences in persuasion may arise from differences in belief updating or from differences in payoffs. The definition of persuasion that we adopt—"influencing behavior via provision of information" (Kamenica, 2019)—applies to both possibilities, and thus we refer to "persuadable agents" without further distinguishing between the two potential underlying mechanisms.

[26] In our pre-registration, we specified that in some specifications, we would pool data from a pilot ($N = 2,019$) with the data from the main experiment. The pilot instrument was virtually identical to the instrument used in the main experiment. We report both unpooled and pooled specifications.

3.2 Experimental design

Figure 1 outlines the structure of Experiment 1. We tell all respondents about a recent study (Lott, 2018) which finds that "undocumented immigrants are at least 142% more likely to be convicted of a crime than other Arizonans ... they also tend to commit more serious crimes and serve 10.5% longer sentences, are more likely to be classified as dangerous, and are 45% more likely to be gang members than U.S. citizens."[27] We also truthfully tell our respondents that a number of sources (including a researcher affiliated with the Cato Institute, a libertarian think tank) have recently challenged some of the study's methods, claiming that errors in analysis invalidate its results.[28]. We then tell participants that we conducted a project on political and social attitudes in the United States earlier in the year, and that respondents to this previous study were given an opportunity to authorize a \$1 donation to Fund the Wall: a potentially stigmatized expression of anti-immigrant beliefs. We inform participants that we have matched them with one of these respondents, and that this respondent chose to authorize the donation. Respondents in the *Excuse* condition are (truthfully) told that their matched respondent was informed about the study before deciding whether or not to authorize the donation to Fund the Wall, while respondents in the *No Excuse* condition are (truthfully) told that their matched respondent was not informed about the study before making their donation decision.

Measuring type inference   After learning whether or not their matched respondent knew about the study, all participants respond to the following open-ended question: "Why do you think your matched respondent chose to donate to Fund the Wall?" As we discuss in Section 3.3, these open-ended responses form the raw data for our first measure of type inference; we employ text analysis to systematically analyze the open-ended responses. Participants are then cross-randomized into one of two conditions: "tolerance" and "gullibility".[29,30] Participants in the "tolerance" condition

---

[27]This study has been widely covered by the media, including *The Washington Times*, *National Review*, and *Fox News*, and has been repeatedly cited by Trump administration officials. For example, in a January 2018 speech on "national security and immigration priorities of the administration," then-Attorney General Jeff Sessions claimed that the study proved that "tens of thousands of crimes have been committed in this country that would never have happened if our immigration laws were enforced and respected like they ought to be" (see footnote 5).

[28]In order to ensure that our respondents are not misinformed, we debrief them at the end of the study and provide them with a meta-analysis summarizing the work on the effects of immigration on crime (Ousey and Kubrin, 2018).

[29]Of course, as described above, gullibility is only one of a set of potential reasons for donating after being exposed to information suggesting immigrants commit more crimes; alternative reasons include lower tolerance for crime, higher levels of risk aversion, etc. We focus on gullibility in our experiment because it is (arguably) the most natural "second type," because it was the most frequent reason cited in our pilot results, and because it is more easily coded.

[30]We measure type inference using a "between" design (in which each respondent is asked only about a single dimension) rather than a "within" design (in which respondents are asked about both dimensions). We employ a

are told that their matched respondent completed the "Foreign Culture Tolerance Scale," a "short questionnaire measuring tolerance toward foreign values and traditions," before making their donation decisions. Participants in the "gullibility" condition are told that their matched respondent completed the "Gullibility Scale," a "short questionnaire which measures how easily people are manipulated by evidence from untrustworthy sources," before making their donation decisions. All participants are asked to guess their respondent's score; we incentivize this guess by informing them that if they correctly guess the score, they will be entered into a lottery for a \$50 Amazon gift card.[31]

3.3 Main results

Empirical strategy   To identify the effect of the excuse on respondents' inference about the matched respondent's type, we estimate the following empirical specification:

$$y_i = \alpha_0 + \alpha_1 \text{Excuse}_i + \varepsilon_i, \tag{3}$$

where Excuse_i is an indicator taking value 1 for participants in the *Excuse* condition and value 0 in the *No Excuse* condition. y_i is our participant's belief about the matched respondent's type. We employ robust standard errors throughout.

Main findings   We begin by using text analysis to measure how participants respond to the open-ended question "Why do you think your matched respondent chose to donate to Fund the Wall?" The advantage of this approach is that we can directly measure what comes to respondents' minds rather than drawing their attention to the particular dimensions we are interested in. Measuring type inference through analyzing open-ended text responses may thus better capture the natural process of inference than directly asking about perceptions of tolerance or gullibility.[32]

---

between design in order to minimize experimenter demand effects and to avoid order effects (Haaland et al., 2020).

[31]The previous study respondents with whom Experiment 2 subjects were matched completed a survey very similar in structure to our Experiment 1 survey, but the two surveys were not precisely the same. In particular, it was important that Experiment 2 subjects believe that their matched respondents completed the scale *before* learning about the Lott study and before making their donation decision, such that subjects' inferences about their matched respondents' scores were not biased by subjects believing that learning about the Lott study changed their matched respondents scores. However, administering these scales in this manner to participants in Experiment 1 might have created significant demand effects, compromising the validity of our findings. To avoid deception, we thus ran a small auxiliary survey before we ran Experiment 2, and we matched Experiment 2 subjects with participants from this auxiliary survey.

[32]Because respondents in both the *No Excuse* and *Excuse* conditions see the same question, our approach also mitigates concerns about experimenter demand. We discuss experimenter demand in greater depth in Section 3.4.

We began with five "seed words" for each type. For (in)tolerance, we chose *racist*, *biased*, *xenophobic*, *intolerant*, and *prejudiced*. For gullibility, we chose *convinced*, *persuaded*, *gullible*, *naive*, and *sucker*. We added all "most relevant" synonyms for these words, as classified by the website www.thesaurus.com. In order to capture different parts of speech, we then stemmed all words in our list (e.g., xenophobic → xenophob, gullible → gullib), for a total of 23 intolerance-related stems and 30 gullibility-related stems (Gentzkow et al., 2019).

We then define two indicator variables — one variable that takes value 1 if the respondent uses an intolerance-related stem and 0 otherwise, and another variable that takes value 1 if the respondent uses a gullibility-related stem and 0 otherwise — and estimate treatment effects on the probability that the respondent uses at least one word in each list.[33] In order to eliminate potential degrees of freedom for analysis, we pre-specified this entire procedure, including the list of stems and the code file used for analysis.

Figure 2 displays results from our text-based type inference. Participants in the *Excuse* condition are 7 percentage points less likely to use a stem related to intolerance when describing their matched respondent's motive, compared to a mean of 17 percent among participants in the *No Excuse* condition ($p < 0.001$). These same participants are also 3 percentage points more likely to use words related to gullibility ($p < 0.001$), relative to a mean of 7 percent among participants in the *No Excuse* condition.[34] These are substantial effect sizes, which highlight that the availability of a rationale strongly changes people's inference about their matched respondent's motives. Columns 1–3 of Table 1 display results in regression form and demonstrate robustness to the inclusion of demographic and partisan controls.

Figure 2 also displays results from our structured belief measures. Participants who believe their matched respondent had an excuse rated their respondent 0.13 standard deviations lower on the intolerance scale ($p < 0.001$), and 0.32 standard deviations higher on the gullibility scale ($p < 0.001$). As with the text analysis measure, effects are similar in the pilot and in the pre-registered main experiment, are robust to the inclusion of control variables, and are precisely estimated. Table 2 displays results in regression form and demonstrates robustness to the inclusion of demographic and partisan controls. To further validate our two measures of type inference, we show in Table B3 that they are highly correlated: on average, a respondent who uses a word related to intolerance

---

[33]These two outcomes are neither mutually exclusive nor jointly exhaustive; responses that contain both an intolerance-related stem and a gullibility-related stem will have both intolerance and gullibility indicators equal to one, whereas responses that contain neither type of stem will have both indicators equal to zero. Thus, our results are unbiased even if participants perceive a nonzero correlation between intolerance and gullibility.

[34]We were intentionally conservative when choosing stem words in order to minimize the rate of false positives.

(gullibility) when describing the matched respondent's motive rates the matched respondent as half a standard deviation more intolerant (gullible) than a respondent who does not use such a word.

As a final measure of beliefs, we use a Support Vector Machine regression to *predict* participants' ratings of their partners on the intolerance and gullibility scales based upon their open-ended text responses. This measure aims to exploit all information given in the open-responses for predicting the subsequent belief measures; importantly, because all participants completed the open-ended responses before seeing the scale, we can include participants assigned to the gullibility condition when evaluating effects on intolerance, and we can include participants assigned to the intolerance condition when evaluating effects on gullibility.[35] We then estimate treatment effects on these *predicted* scores. Unlike our word-counting exercise, this approach is insensitive to the set of keywords chosen and is thus a more disciplined procedure to measure perceptions of bias and gullibility expressed in participants' open-ended responses. Columns 4–6 of Table 1 display these results in regression form and confirm that our treatment has a statistically and economically significant effect on perceptions.

Taken together, our evidence suggests that when judging others' motives, people believe that those who donated with an excuse are more persuadable and less intolerant than those who donated without an excuse.

3.4   Ruling out alternative explanations

Demand effects   One potential concern regarding the validity of our estimated treatment effects is that respondents across different treatment conditions hold different beliefs about the experimenter's expectations, and that these beliefs drive our findings. Despite recent evidence that respondents are not elastic to explicit signals of the experimenter's expectations in online surveys (de Quidt et al., 2018), suggesting a limited quantitative importance of demand effects in the context of our experiment, we provide direct evidence on beliefs about the purpose of the study.

We measured respondents' beliefs about the purpose of the experiment at the end of Experiment 1 using an open-ended question: "If you had to guess, what was the purpose of this study?" To examine whether respondents in the different treatment conditions hold different beliefs about the purpose of the study, we employ machine learning techniques to classify these text responses. In particular, a Support Vector Machine classifier trained to predict treatment status given the

---

[35]We employed a radial basis function kernel in the SVR, though in practice the results are insensitive to other choices of parameters. This final exercise was not pre-registered.

participant's response cannot predict whether respondents were assigned to the *Excuse* or *No Excuse* condition better than chance (Table B4).[36] This suggests that the treatment does not significantly affect respondents' perceptions about the purpose of the study.[37]

Differential attrition Could differential attrition across treatment arms explain our findings? Attrition rates in Experiment 1 are virtually identical among respondents in the *Excuse* and *No Excuse* conditions ($p = 0.23$) and neither political affiliation nor any other demographic variable systematically predicts differential attrition across treatment arms (Table **??** in the Appendix).

3.5 Do people dislike the intolerant more than the gullible?

A key assumption in our model is that the social image cost of being perceived as intolerant is higher than the social image cost of being perceived as gullible.[38] In an additional auxiliary experiment, we confirm that this assumption is reasonable.[39] In particular, we recruit a representative sample of Democrats (the same population as in Experiment 1), and we inform them that they have been matched either with another respondent who scored above average on The Gullibility Scale (gullibility condition) or below average on The Foreign Culture Tolerance Scale (intolerance condition). We then give them the opportunity to authorize a $1 donation to their matched respondent. Consistent with our assumption, we observe significantly higher donations to matched partners with high gullibility than to matched partners with high intolerance (see Appendix Figure B1 and Appendix Table B6).

4 Excuses and Xenophobic Expression

Experiment 1 demonstrated that subjects were more likely to ascribe xenophobic behavior to gullibility and less likely to ascribe xenophobic behavior to intolerance when an excuse was available. Furthermore, the auxiliary experiment demonstrated that subjects were less spiteful toward gullible types than intolerant types. These results raise the question of whether people who hold intolerant

---

[36]We employed a radial basis function kernel in the SVR, though in practice the results are insensitive to other choices of parameters.

[37]In Section 4.2, we present evidence that our classifier is sufficiently sensitive to detect differences in stated beliefs when they are present.

[38]For simplicity, we assume in our model that there is no direct social image cost of being perceived as gullible. We can relax this assumption without changing any of the key results, so long as the direct social image cost of being perceived as intolerant remains higher than the direct social image cost of being perceived as gullible.

[39]A copy of the experimental instructions is available in Appendix Section H.

views strategically use excuses to lower the social cost of publicly expressing these views (Proposition 2 of our theoretical framework). We examine this question in Experiment 2 in the same context as in Experiment 1: anti-immigrant expression justified by claims that immigrants disproportionately commit violent crimes.

4.1 Experiment 2: sample and experimental design

We again worked with Luc.id to recruit 3,728 self-reported Republicans and Independents. In some specifications, we supplement this data with approximately 716 Republicans and Independents from a pilot experiment with Luc.id, also conducted in January 2020, that had nearly identical wording.[40] In other specifications, we additionally pool with data from a replication and robustness check (1373 Republicans and Independents) conducted in October 2020.[41] Participants were directed to our survey on the online platform *Qualtrics*; only participants who were over the age of 18, resided in the United States, indicated their consent to participate, and passed a simple test of attention were allowed to proceed.[42] Our sample of respondents is broadly representative of Independents and Republicans in the United States (Appendix Table B7) and is well-balanced on observables across treatment arms (Appendix Table B8). We pre-registered all experimental procedures and analyses for the main experiment. Figure 3 outlines the structure of Experiment 2.

Information: Lott study   After completing a series of demographic questions, respondents are assigned to one of three different treatment conditions: an *Excuse* condition, a *No Excuse* condition, and a *Control* condition. Respondents in the *Excuse* and *No Excuse* conditions receive information about the Lott study; respondents in the *Control* condition do not learn about the study.

Donation decisions   To minimize experimenter demand concerns, we truthfully tell our respondents that we will randomly select one of two organizations — an anti-immigrant organization (*Fund the Wall*) and a pro-immigration organization (*Texas Civil Rights Project*) — and the respondents will have the opportunity to authorize a $1 donation to this organization.[43] In practice, we randomized almost all respondents to Fund the Wall to maximize statistical power for our comparison of interest.

---

[40]We pre-registered reporting both results pooling pilot data with our main data and results with our main data alone.

[41]We describe this replication in greater detail in Section 4.2.

[42]All survey instruments are available in Appendix E.

[43]Respondents are explicitly informed that their own survey reward will be unaffected by their choice of whether or not to donate: they are simply authorizing the researchers to make a donation on their behalf.

Visibility manipulation   Our treatment hinges upon respondents' decisions being publicly observable. We ask respondents to consent to us accessing their name, city, and operating system from the survey provider (which confirmed that they would provide us with this data subject to participant consent) and give respondents the option to terminate the survey if they do not consent.[44] We inform respondents that once the Lott study is published in a reputable academic journal, we will post the results from the survey, including their individual donation decision, on our study website.[45]

We also inform our respondents that "As researchers, we believe it is important to communicate our findings about political and social attitudes in [City of respondent] to the public".[46] We then inform our respondents that "we will work with major news organizations in [City of respondent] with both a liberal and conservative viewership to publicize our website through newspaper and website articles", and "we will also promote our website via Facebook ads to [City of respondent] residents". This non-deceptively generates a plausible social cost for acting in a way that will be stigmatized in the respondent's area.[47]

Varying the availability of the excuse   Our main object of interest is to identify the excuse effect. This is complicated by the fact that providing information to respondents may affect their behavior through two alternative channels other than the availability of the excuse. First, the information might be directly persuasive, leading more respondents to donate because their private views have changed. Second, even if the information does not persuade respondents, respondents might believe that their *audience will be persuaded* by the study's description on the website, leading respondents to expect lower social stigma from donating and thus increasing donation rates. We thus design our experiment to rule out these competing effects. To hold fixed the first mechanism, all respondents in the *Excuse* and *No Excuse* condition receive the same information about the study. To hold fixed the second mechanism, we show respondents in the *Excuse* and *No Excuse* conditions screenshots of our website, clearly indicating that all website visitors will be informed about the study.

---

[44]Because participants consent before exposure to the treatment, any resulting attrition should not bias our estimates within the population who completes the survey.

[45]At the time of writing (October 2020), the Lott study has not yet been published, and thus the donation decisions are not yet online.

[46]We used participants' IP address to capture and display their current location (i.e. their city). The IP addresses were subsequently deleted to protect respondents' privacy.

[47]Once the study has been published and the website has been created, we plan to publicize the website through Facebook Advertisements and by disseminating the results of our study to news organizations. In line with instructions to respondents, the website will only be created once the Lott study has been published.

The key experimental treatment thus cleanly varies the availability of an excuse for donating. In the *Excuse* condition, we inform respondents that "Website visitors will know that you knew about the results of Dr. Lott's study," giving respondents an excuse to donate (i.e. believing, based on the findings of the Lott study, that illegal immigrants commit substantially more crime than US citizens). Respondents also see a screenshot of the website, which clearly states that "All participants were told about Dr. Lott's study." Thus, respondents in the *Excuse* condition expect that their audience will know they learned about the study before donating.

Conceptually, in the *No Excuse* condition, we would like to show respondents a website screenshot stating that "No participants were told about Dr. Lott's study". However, because these participants did in fact learn about the study, such a screenshot would be deceptive. Instead, we exploit the fact that Lott's study had not yet been published in an academic journal (a fact about which we explicitly informed all respondents when describing the website). In particular, we show respondents a website screenshot stating that "We surveyed respondents earlier this year before Dr. Lott's study was published." In the survey, we write that "the website states that you were surveyed before the study was published and does not mention that you were shown an early summary of the study's findings."[48] Respondents in this condition thus believe that their audience will believe that they (the respondents) had no information excusing their decision to donate to Fund the Wall.

Control condition We also include a *Control* condition in which neither the respondent nor the audience learns about the Lott study. This condition allows us to estimate the combined effects of direct persuasion and anticipated persuasion of the audience, as we describe below.

4.2   Experiment 2: main results

Empirical strategy To identify the joint effects of direct persuasion and anticipated persuasion of the audience (i.e. the direct persuasive effect of learning about the Lott study in addition to the indirect effect of learning that one's audience has learned about the Lott study and may thus be more likely to approve of the donation), we compare the *Control* condition with the *No Excuse* condition. To identify the excuse effect, we compare the *No Excuse* condition to the *Excuse* condition. This design thus allows us to benchmark the excuse effect against the combined effect of first- and second-order persuasion. Our main specification of interest is given as follows:

---

[48]As discussed in Section 4.2, we replicate our findings using a different and more subtle treatment manipulation.

$$y_i = \beta_0 + \beta_1 \text{Excuse}_i + \beta_2 \text{Control}_i + \varepsilon_i \tag{4}$$

where y_i is an indicator taking value 1 if the respondent authorized the donation to Fund the Wall and 0 otherwise; Excuse_i is an indicator taking value 1 if the respondent was assigned to the *Excuse* condition and 0 otherwise; and Control_i is an indicator taking value 1 if the respondent was assigned to the *Control* condition and 0 otherwise. The omitted category is thus the *No Excuse* condition. We employ robust standard errors throughout our analysis.

Main findings   Table 3 and Figure 4 display the main findings of Experiment 2. We find a large and statistically significant effect on respondents' willingness to authorize a donation to Fund the Wall: respondents in the *Excuse* condition are 6.3 percentage points more likely to authorize the donation than respondents in the *No Excuse* condition. This effect is highly statistically significant ($p < 0.001$) and large relative to a control condition mean of 48.8 percentage points. Effect sizes are almost identical in our pre-specified main study, our pilot study, and a replication several months later. The estimated effects are also stable to the inclusion of demographic controls; Figure B2 presents a "coefficient stability plot" (Rao, 2020) displaying coefficient estimates under every possible combination of controls.

In contrast to the *Excuse* vs. *No Excuse* comparison, respondents in the *No Excuse* condition are only 0.007 percentage points more likely to authorize a donation than respondents in the control condition, suggesting that the combined effects of first- and second-order persuasion are small. Relatively small persuasion effects are in line with other information provision experiments in the immigration domain, which typically find relatively small or null effects on behavior and stated preferences (Hopkins et al., 2019; Alesina et al., 2019; Grigorieff et al., 2020). Thus, small effects of anticipated persuasion are consistent with agents holding accurate expectations about whether their audience will be persuaded.

Given the small joint effect of persuasion and the anticipated persuasion of the audience, what might explain the large excuse effect we observe? First, agents may simply hold incorrect higher-order beliefs: in particular, they may believe that their audience is more likely to believe that they have been persuaded by the information. Alternatively, they may predict that social rewards or sanctions associated with being perceived as intolerant are not linear in the probability that one is intolerant: for example, they may believe that as long as it appears that there is some small

probability that they are not intolerant (i.e. because they were exposed to the study and may have been persuaded), their audience will refrain from socially sanctioning them ("innocent until proven guilty"). To preserve analytic tractability and convey our intuition as simply as possible, we do not formally model either of these channels.

Heterogeneity by local vote shares   An implication of our model is that the audience's composition — the share of tolerant vs. intolerant agents — should affect donation decisions by changing the perceived judgment associated with donating. Because we informed respondents that we would promote the website (on which their individual donation decision would be posted) within their geographical area, we might expect that controlling for the respondents' own private views, respondents in areas with a greater fraction of Republicans should be less sensitive to the availability of a rationale than respondents in areas with a lower fraction of Republicans, since Republicans are likely to approve of the decision to donate to Fund the Wall even in the absence of a rationale.[49] We thus pre-registered investigating heterogeneity by the 2016 Republican vote share of respondents' county, which we do by estimating the following specification:

$$
\begin{aligned}
y_i = \beta_0 + \beta_1 \text{Excuse}_i + \beta_2 \text{Control}_i + \beta_3 \text{Excuse}_i \times \text{Rep share}_i + \beta_4 \text{Control}_i \times \text{Rep share}_i \\
+ \beta_5 \text{Rep share}_i + \varepsilon_i
\end{aligned}
\tag{5}
$$

Table 4 displays the results, revealing striking heterogeneity by the Republican vote share of respondents' counties. Although the heterogeneity is only statistically significant when we control for individuals' demographic characteristics, the effect is large in magnitude: a one standard deviation increase in the Republican vote share of a respondent's county is associated with halving the magnitude of the excuse effect. These results should not be interpreted as a causal effect of respondents' audiences — it may be, for example, that Republicans in Democratic areas feel greater need to signal their support for the study by publicly donating — but the pattern is consistent with our intuition that the excuse effect should be larger when the share of agents who privately approve of the action is smaller.

Robustness experiment   The excuse effect, though potentially large in magnitude, is difficult to cleanly identify given the need to hold both first-order and second-order persuasion constant. The instructions of Experiment 2 were thus relatively explicit: participants were directly informed about

---

[49]Of course, it is possible that Republicans care less about the opinions of Democrats than that of fellow Republicans, which would weaken the heterogeneity results we examine.

their audience's information sets and were reminded about this information when making their decision, thus ensuring that they fully understood the instructions. Yet one concern is that these instructions, by making the higher-order belief mechanism salient, induced experimenter demand effects that biased our estimated treatment effects. In October 2020, we conducted an additional robustness experiment with 1,373 Republicans, which used a much leaner set of instructions and which made the issue of higher-order beliefs much less salient.[50] In this experiment, we did not show participants any screenshots or illustrations; we did not explicitly tell them what their audience was likely to believe about whether they had seen the study prior to making the donation; and we considerably shortened the survey. We omitted the *Control* condition, given our primary interest in the comparison between Excuse and No Excuse.

Columns 4–6 of Table 3 report our estimates. We find similar and statistically significant treatment effects (though of slightly smaller magnitudes) using this much leaner set of experimental instructions. Thus, the results of this robustness experiment not only mitigate concerns that experimenter demand effects drove our original findings but they also bolster their external validity: despite being conducted almost a year later in a substantially different economic and political situation (in the midst of the COVID-19 pandemic and an economic recession, just a few weeks before Election Day), we once again find significant evidence for the excuse mechanism. We also replicate the heterogeneity by Republican vote share (Columns 4–6 of Table 4), which, if anything, is stronger than in the original experiment.

Demand effects: further discussion   *A priori*, it seems plausible that showing participants information about the Lott study may induce demand effects and thus affect donation rates. Even if such demand effects are present, however, they will not bias our main comparison of interest (*Excuse* vs. *No Excuse*), given that participants in both treatment arms are shown identical information about the study. Another concern is that the treatment manipulation of beliefs about the audience's information set induced differential experimenter demand effects. In addition to conducting the robustness experiment described above, we address this concern through a number of additional exercises.

As in Experiment 1, we measured respondents' beliefs about the purpose of the experiment at the end of Experiment 2 using an open-ended question. As before, we use a Support Vector

---

[50]We targeted 1,426 complete responses, the sample size needed for 80% power to detect an effect size of the same magnitude as the effect size estimated in the main experiment and pilot (Column 3 of Table 3). Attrition rates were slightly greater than in previous rounds, resulting in a slightly smaller sample. See Appendix F for the experimental instructions.

Machine classifier to predict treatment status given the participant's response. Employing 75 percent of our sample as a training set and the remaining 25 percent as a test set, we show that we cannot predict treatment status better than chance when distinguishing between the *Excuse* and *No Excuse* conditions (Table B11). However, we can predict assignment to the control condition substantially better than chance (Table B11), which highlights that respondents in the control condition hold different beliefs from respondents in the *Excuse* and *No Excuse* condition. Given that the control condition differs significantly from the *Excuse* and *No Excuse* conditions in that control respondents do not learn about the Lott study, this difference is to be expected; we view this result as validation for our method, as it demonstrates that we would in principle detect differences in perceived purpose between the *Excuse* and *No Excuse* conditions if such differences were present.

In addition to the machine-learning exercise, we also hired two independent research assistants to hand-code the responses to the open-ended purpose question. Table B10 in the Appendix shows that the majority of our respondents believed that we wanted to study the effects of information on anti-immigrant sentiment or participant's willingness to have their decisions posted on the website. Fewer than 1 percent of our sample correctly guessed the true purpose of our experiment (Column 1). Table B10 also shows that on almost all of the dimensions we code, beliefs about the purpose of the study do not significantly differ between the *Excuse* and *No Excuse* conditions. The exception is Social Image (Column 3): respondents in the *Excuse* condition are 2 percentage points more likely than respondents in the *No Excuse* condition to believe that the study was about whether people were willing to publicly express political views ($p = 0.038$). Although statistically significant, this difference is small in magnitude and cannot explain our effect sizes. Reassuringly, respondents were no more likely to believe that the experimenters were biased in the *Excuse* condition than in the *No Excuse* condition (Column 6, $p = 0.994$). As suggested by the results of the machine learning exercise described previously, we do find significant differences in perceived purpose between the control condition and the *No Excuse* condition and also between the control condition and the *Excuse* condition. This is likely due to the fact that we provided respondents in the *No Excuse* and *Excuse* conditions information suggesting that undocumented immigrants commit more crimes than US citizens (i.e. the Lott study), while we did not provide any such information to respondents in the control condition. However, these differences do not affect our main comparison of interest (*No Excuse* vs. *Excuse*).

Finally, heterogeneous treatment effects by the county-level Republican vote share provide additional suggestive evidence against experimenter demand effects driving our findings. In particular,

for demand effects to explain our treatment effects, we would require that — controlling for respondents' own characteristics, including their political affiliation — respondents in counties with a lower Republican vote share are substantially more affected by latent experimenter demand effects (that both our hand-coding and our machine learning exercise fail to detect, and which are induced even by the leaner set of experimental instructions used for the robustness experiment) than respondents in counties with a higher Republican vote share. While not impossible, we view this contingency as unlikely.

Differential attrition Could patterns of differential attrition explain the estimated treatment effects in our data? We find no differential attrition among respondents in the Excuse versus *No Excuse* condition ($p = 0.47$), and there is no evidence of differential attrition between different subgroups (Table B12 in the Appendix). We do find a precisely estimated four percentage point lower attrition rate among respondents in the control condition compared to respondents in the *Excuse* condition and the *No Excuse* condition ($p < 0.001$), which may be explained by the greater survey length of the *Excuse* and *No Excuse* versions of the survey. This does not affect our estimates of the main effect of interest (*No Excuse* vs. *Excuse*), but may slightly bias the benchmark (*Control* vs. *No Excuse*).

5   Policy Implications and Conclusion

Motivated by a global wave of anti-immigrant rhetoric and policy, we study how commonly known rationales to oppose immigration serve as excuses to justify anti-immigrant behavior. In a motivating survey, we find that public support among Republicans for a permanent ban on Mexican immigration into the US jumps from 32% to 51% if respondents are allowed to attach a justification for their public support: protecting the US from contagious diseases. We then use large-scale experiments to examine the mechanisms through which excuses facilitate the expression of anti-immigration behavior, focusing on one of the most widely-cited justifications for reducing immigration: the claim that immigrants commit crimes at vastly higher rates than citizens. In a first experiment, we show that subjects perceive donors who had been exposed to an anti-immigrant rationale as less biased against immigrants and more persuadable than donors who had not been exposed. In a second experiment, we show that subjects who believe that their exposure to the rationale will be publicly observable are substantially more likely to make the donation to an anti-immigrant organization than subjects who believe that their exposure to the rationale will remain

private.

Our approach can be applied to understand a variety of political economy phenomena. For example, populist rhetoric often seeks to generate common knowledge — or the perception of common knowledge — of excuses. Müller (2016) argues that populist rhetoric is often characterized by appeals to the beliefs or desires of the "people" or a "silent majority" — a group which often has little to no basis in fact. For example, several commentators have highlighted Donald Trump's tendency to use phrases such as "People say ..." when discussing politically sensitive issues, and as Rosenblum and Muirhead (2019) argue, this practice is common to a number of prominent populist politicians around the world spanning the ideological spectrum. Such rhetoric generates the perception of *common knowledge* of the excuse: by implying that fringe conspiracy theories are known to a large group of people (and by appearing to endorse the theory themselves), populists seek to convey that the excuse will be credible and thus effective.[51] Closely related is the tendency to rely on anecdotes — for instance, isolated cases of voter fraud — in order to argue certain phenomena are widespread, ignoring statistics to the contrary. Also related is dog-whistling: "sending a message to certain potential supporters in such a way as to make it inaudible to others whom it might alienate or deniable for still others who would find any explicit appeal along those lines offensive" (Goodin and Saward, 2005), which has been used to describe the Republican Party's "Southern Strategy" to win white support in the South by appealing to racial tensions (Haney-López, 2014).[52] As with "people say" and related language, "dog-whistles" generate two types of excuses: one for the politician vis-a-vis the public, and one for the politician's supporters vis-a-vis others who disapprove of the statement, allowing them to publicly support the politician and his or her policies without incurring social stigma.[53]

Our findings are also relevant for the debate about the influence of fake and misleading news on society. While studies suggest that the persuasive effect of fake news is limited (Nyhan, 2018; Allcott and Gentzkow, 2017), our study points to an alternative mechanism through which fake news can affect public expression. In particular, fake news can generate a "persuasion multiplier":

---

[51]This practice is, of course, also consistent with populists exploiting social learning channels in order to bolster the persuasive power of their claims.

[52]In a 1981 interview, Republican strategist and Republican National Committee chairman Lee Atwater described the strategy as follows: "You start out in 1954 by saying, 'N—, n—, n—.' By 1968 you can't say 'n—': that hurts you. Backfires. So you say stuff like forced busing, states' rights and all that stuff. You're getting so abstract now [that] you're talking about cutting taxes, and all these things you're talking about are totally economic things and a byproduct of them is [that] blacks get hurt worse than whites. And subconsciously maybe that is part of it. I'm not saying that." (Lamis, ed, 1999)

[53]Indeed, a third type of excuse may be a "self-excuse" for politician's supporters who do not want to admit to themselves that they endorse racist positions, as in Bénabou and Tirole (2011a).

rationales that plausibly persuade a small subset of the population and are commonly known to exist can change public behavior among a much larger fraction of the population, increasing their willingness to express otherwise-stigmatized views by increasing the effectiveness of their excuse.

This insight has implications for debunking fake news spread online and offline. In particular, our findings suggest that in order to prevent a given fake news story from spreading, it might be insufficient to debunk it *privately*; instead, it is crucial to generate common knowledge that the excuse is invalid. This insight has valuable implications for institutional policy. Among other platforms, Facebook has experimented with various strategies to curtail the spread of misinformation, including warning users before they post an article flagged as fake news and flagging fake or misleading news when it appears on users' timelines (e.g., because a friend shared it). The former initiative maps closely onto a "first-order" debunking in our model (private persuasion), while the second initiative maps onto a "second-order" debunking (debunking one's audience). Yet to the extent that Facebook does not yet debunk all users (more precisely, to the extent that the fact that Facebook does not debunk all users is not common knowledge), it generates a ready-made excuse for sharing fake news: posters can credibly claim that they were not warned the news was fake.[54] Our results suggest it is important not only to debunk both the poster and the audience, but also to make it clear to the poster that the audience will *know that he or she was debunked before posting*. This could be done by including a screenshot in the warning shown to the poster of what his or her post will look like to others, in which the sentence "The poster was warned that this link has been flagged as fake or misleading before posting" is clearly visible. An alternative and simpler path would be to simply roll out the feature to the entire user-base, generating common knowledge that all users are warned before posting fake news. Because the general equilibrium results of such a change differ significantly from the partial equilibrium results by creating common knowledge, current estimates of the effects of debunking on users' propensity to share fake news may substantially understate the true effects that would be realized if platforms were to fully scale up the feature.

Our results suggest several directions for further research. First, what implications do our results have for the "supply side" of excuses: can "excuse entrepreneurs" who are able to generate common knowledge about plausible rationales to act in a potentially stigmatized manner cause striking reversals of social norms, even if their persuasive impact is limited, and can similar patterns

---

[54]Indeed, Facebook's fact-checking efforts have been widely criticized for a lack of transparency, and it is thus certain that most Facebook users lack information about how the platform fights misinformation. (Nyhan, Brendan. "Why the Fact-Checking at Facebook Needs to Be Checked." *The New York Times*, October 23, 2017.)

help explain the rising popularity of ideologically extreme media outlets? Moreover, can growing partisan polarization in media consumption make excuses more effective by allowing partisans to more credibly claim that they have not been exposed to information contradicting their views?

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya, "Radio and the Rise of The Nazis in Prewar Germany," *The Quarterly Journal of Economics*, 2015, *130* (4), 1885–1939.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva, "Immigration and redistribution," Working Paper 24733, National Bureau of Economic Research, 2019.

Ali, S. Nageeb and Roland Bénabou, "Image versus Information: Changing Societal Norms and Optimal Privacy," *American Economic Journal: Microeconomics*, 2020, *12* (3), 116–164.

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

Andreoni, James and B. Douglas Bernheim, "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 2009, *77* (5), 1607–1636.

Ariely, Dan, Anat Bracha, and Stephan Meier, "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 2009, *99* (1), 544–555.

Arrow, Kenneth, "The Theory of Discrimination," in Albert Rees and Orley Ashenfelter, eds., *Discrimination in Labor Markets*, Princeton, New Jersey: Princeton University Press, 1973.

Bandura, Albert, Claudio Barbaranelli, Gian Vittorio Caprara, and Concetta Pastorelli, "Mechanisms of Moral Disengagement in the Exercise of Moral Agency," *Journal of Personality and Social Psychology*, 1996, *71* (2), 364–374.

_ , Gian Vittorio Caprara, Claudio Barbaranelli, Concetta Pastorelli, and Camillo Regalia, "Sociocognitive Self-regulatory Mechanisms Governing Transgressive Behavior," *Journal of Personality and Social Psychology*, 2001, *80* (1), 125–135.

Becker, Gary S., *The Economics of Discrimination*, University of Chicago Press, 1957.

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

_ **and** _ , "Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

Benabou, Roland and Jean Tirole, "Laws and Norms," Technical Report w17579, National Bureau of Economic Research 2011.

Bénabou, Roland, Armin Falk, and Jean Tirole, "Narratives, Imperatives, and Moral Reasoning," Working Paper 24798, National Bureau of Economic Research, 2018.

Besley, Timothy, Anders Jensen, and Torsten Persson, "Norms, Enforcement, and Tax Evasion," Technical Report, National Bureau of Economic Research 2019.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, *110* (10), 2997–3029.

_ **and Robert Jensen**, "How Does Peer Pressure Affect Educational Investments?," *Quarterly Journal of Economics*, 2015, *130* (3), 1329–1367.

_ , **Georgy Egorov, and Stefano Fiorin**, "From Extreme to Mainstream: The Erosion of Social Norms," *American Economic Review*, 2020, *110* (11), 3522–48.

_ , **Thomas Fujiwara, and Amanda Pallais**, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, *107* (11), 3288–3319.

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

Cohen, Andrew, "How Voter ID Laws Are Being Used to Disenfranchise Minorities and the Poor," 2012.

Cunningham, Tom and Jonathan de Quidt, "Implicit Preferences Inferred from Choice," *Available at SSRN 2709914*, 2016.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang, "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," *Economic Theory*, 2007, *33* (1), 67–80.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

DellaVigna, Stefano and Eliana La Ferrara, "Chapter 19 - Economic and Social Impacts of the Media," in Simon P. Anderson, Joel Waldfogel, and David Strömberg, eds., *Handbook of Media Economics*, Vol. 1 of *Handbook of Media Economics*, North-Holland, 2015, pp. 723–768.

_ , **John A. List, Ulrike Malmendier, and Gautam Rao**, "Voting to Tell Others," *The Review of Economic Studies*, 2017, *84* (1), 143–181.

Enikolopov, Ruben and Maria Petrova, "Chapter 17 - Media Capture: Empirical Evidence," in Simon P. Anderson, Joel Waldfogel, and David Strömberg, eds., *Handbook of Media Economics*, Vol. 1 of *Handbook of Media Economics*, North-Holland, 2015, pp. 687–700.

Ewers, Mara and Florian Zimmermann, "Image and Misreporting," *Journal of the European Economic Association*, 2015, *13* (2), 363–380.

Exley, Christine L., "Excusing Selfishness in Charitable Giving: The Role of Risk," *The Review of Economic Studies*, 2016, *83* (2), 587–628.

Foerster, Manuel and Joel J van der Weele, "Persuasion, Justification and the Communication of Social Impact," 2018.

Frank, Nathaniel, *Unfriendly Fire: How the Gay Ban Undermines the Military and Weakens America*, Macmillan, 2009.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, 2019, *57* (3), 535–74.

Golman, Russell, "Acceptable Discourse: Social Norms of Beliefs and Opinions," *Working paper*, 2020.

_ , **David Hagmann, and George Loewenstein**, "Information Avoidance," *Journal of Economic Literature*, 2017, *55* (1), 96–135.

_ , **George Loewenstein, Karl Ove Moene, and Luca Zarri**, "The Preference for Belief Consonance," *Journal of Economic Perspectives*, 2016, *30* (3), 165–88.

Gonzalez-Barrera, Ana and Philip Connor, "Around the World, More Say Immigrants Are a Strength Than a Burden," *Pew Research Center's Global Attitudes Project*, 2019.

Goodin, Robert E. and Michael Saward, "Dog Whistles and Democratic Mandates," *The Political Quarterly*, 2005, *76* (4), 471–476.

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal, "Does Information Change Attitudes Toward Immigrants?," *Demography*, 2020, *57* (3), 1–27.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," Working paper 8406, CESifo 2020.

Hamman, John R., George Loewenstein, and Roberto A. Weber, "Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship," *American Economic Review*, 2010, *100* (4), 1826–1846.

Haney-López, Ian, *Dog Whistle Politics: How Coded Racial Appeals Have Reinvented Racism and Wrecked the Middle Class*, Oxford ; New York: Oxford University Press, 2014.

Hopkins, Daniel J, John Sides, and Jack Citrin, "The Muted Consequences of Correct Information about Immigration," *The Journal of Politics*, 2019, *81* (1), 315–320.

Jia, Ruixue and Torsten Persson, "Individual vs. Social Motives in Identity Choice: Theory and Evidence from China," Technical Report, National Bureau of Economic Research 2019.

Kamenica, Emir, "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 2019, *11* (1), 249–272.

Kuran, Timur, *Private Truths, Public Lies: The Social Consequences of Preference Falsification*, Cambridge, Massachusetts: Harvard University Press, 1997.

Lacetera, Nicola and Mario Macis, "Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme," *Journal of Economic Behavior and Organization*, 2010, *76* (2), 225–237.

Lamis, Alexander P., ed., *Southern Politics in the 1990s*, Baton Rouge: Louisiana State University Press, 1999.

Langer, Ellen J., Arthur Blank, and Benzion Chanowitz, "The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction.," *Journal of Personality and Social Psychology*, 1978, *36* (6), 635–642.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber, "Sorting in Experiments with Application to Social Preferences," *American Economic Journal: Applied Economics*, 2012, *4* (1), 136–163.

Lott, John R., "Undocumented Immigrants, U.S. Citizens, and Convicted Criminals in Arizona," Working Paper, Social Science Research Network, 2018.

Mendelberg, Tali, *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*, Princeton, N.J: Princeton University Press, 2001. OCLC: ocm46737493.

Müller, Jan-Werner, *What Is Populism?*, Philadelphia: University of Pennsylvania Press, 2016.

Nowrasteh, Alex, "The Fatal Flaw in John R. Lott Jr.'s Study on Illegal Immigrant Crime in Arizona," *Cato Institute*, 2018.

Nyhan, Brendan, "Fake News and Bots May Be Worrisome, but Their Political Power Is Overblown," *The New York Times*, 2018.

Ousey, Graham C. and Charis E. Kubrin, "Immigration and Crime: Assessing a Contentious Issue," *Annual Review of Criminology*, 2018, *1* (1), 63–84.

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

Perez-Truglia, Ricardo and Ugo Troiano, "Shaming Tax Delinquents," *Journal of Public Economics*, 2018, *167*, 120–137.

Pew, "Most Border Wall Opponents, Supporters Say Shutdown Concessions are Unacceptable," *Pew Research Center's Report*, 2019.

Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, 1972, *62* (4), 659–661.

Rao, Aakaash, "Starbility: Coefficient Stability Plots in R (v0.9)," Technical Report, Harvard University 2020.

Rosenblum, Nancy L. and Russell Muirhead, *A Lot of People Are Saying: The New Conspiracism and the Assault on Democracy*, Princeton, New Jersey: Princeton University Press, 2019.

Saccardo, Silvia and Marta Serra-Garcia, "Cognitive Flexibility or Moral Commitment? Evidence of Anticipated Belief Distortion," 2020.

Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal, "Self-Serving Justifications: Doing Wrong and Feeling Moral," *Current Directions in Psychological Science*, 2015, *24* (2), 125–130.

Snow, Mary, "National (US) Poll - July 30, 2019," *Quinnipiac University Poll*, 2019.

United Nations, "UN Expert Denounces the Propagation of Hate Speech Through Social Media," Technical Report, United Nations 2020.

Wines, Michael and Alan Blinder, "Federal Appeals Court Strikes Down North Carolina Voter ID Requirement," *The New York Times*, 2016.

Wood, Thomas and Ethan Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior*, 2019, *41* (1), 135–163.

Yanagizawa-Drott, David, "Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, 2014, *129* (4), 1947–1994.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political Effects of the Internet and Social Media," *Annual Review of Economics*, forthcoming.

Figures
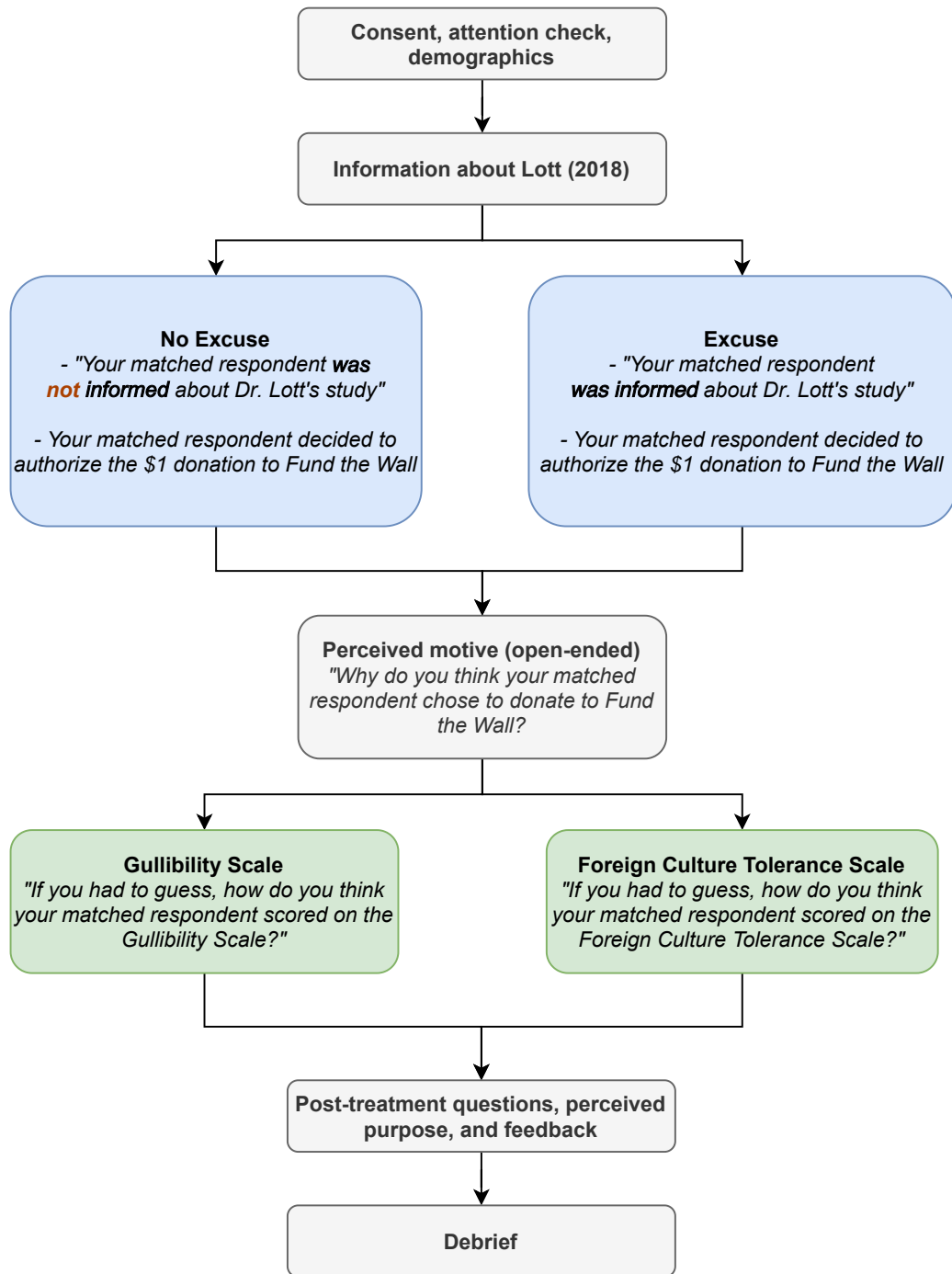
Figure 2: Experiment 1: Type inference based on text analysis and scales

(a) Gullibility: word count

(b) Bias: word count

(c) Gullibility: scale

(d) Bias: scale
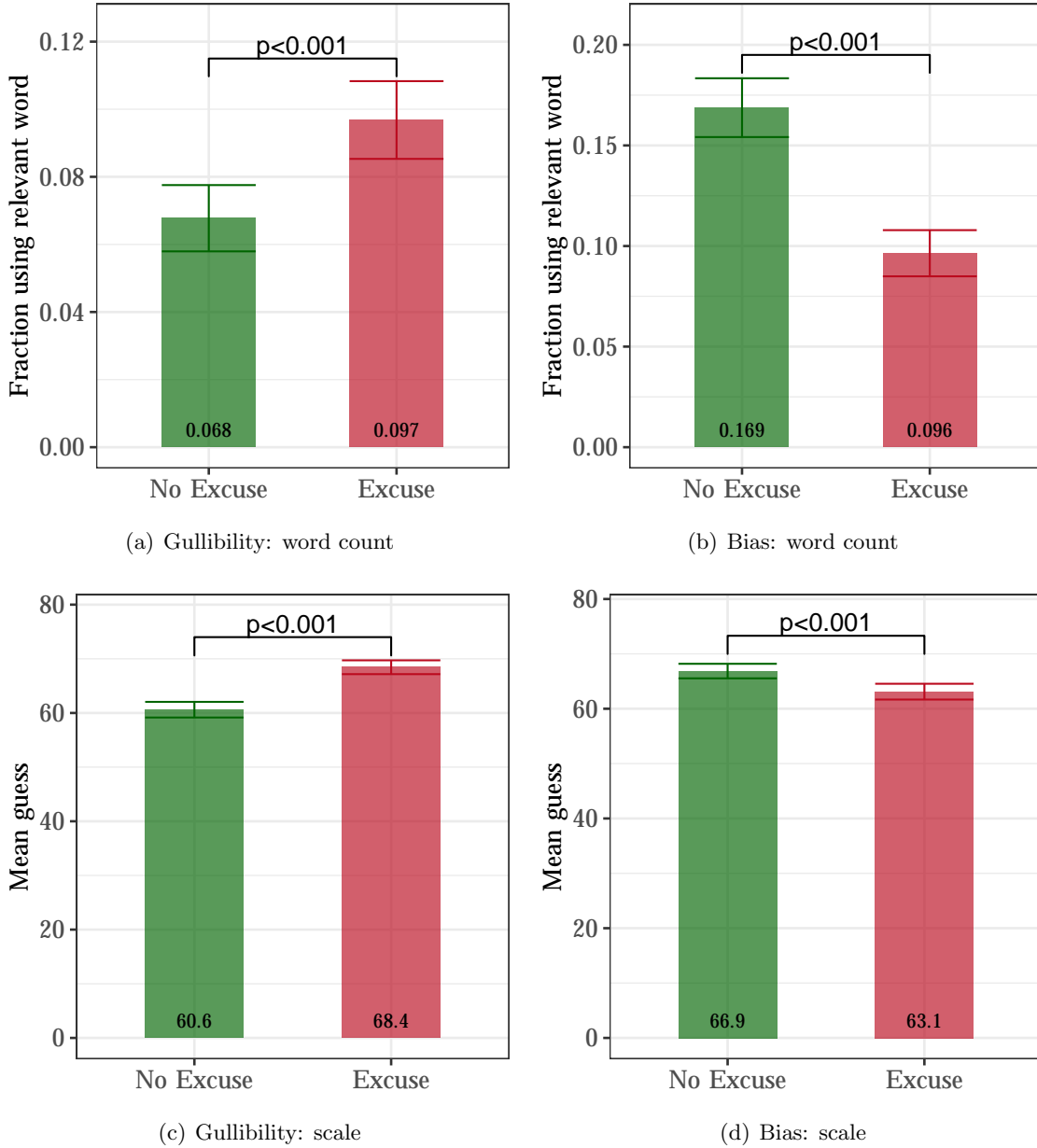
Notes: Figure 2 displays the results from Experiment 2, conducted in February 2020 with a sample of 3,047 Democrats. Panel (a) shows the fraction of respondents who used words related to gullibility across the 'No Excuse' and the 'Excuse' condition. Panel (b) shows the fraction of respondents who used words related to intolerance across the 'No Excuse' and the 'Excuse' condition. Panel (c) shows the mean guess of the matched respondent's score on the Gullibility Scale across the 'No Excuse' and the 'Excuse' condition. Panel (d) shows the mean guess of the matched respondent's score on the (negative of the) Foreign Culture Tolerance Scale scale across the 'No Excuse' and the 'Excuse' condition. The figure displays 95 percent confidence intervals as well as p-values for tests of equality of means across the conditions.

```
                    ┌─────────────────────────────┐
                    │   Consent, attention check, │
                    │        demographics         │
                    └─────────────────────────────┘
```

No Excuse	Excuse	Control

Information about Lott (2018)

Reconsent
"I consent to researchers accessing...first and last name, city, and operating system."

Description of public donation decision and website screenshot
"The page lists individual donation decisions and whether each participant decided to authorize the donation to Fund the Wall"

No Excuse: Excuse manipulation *"The page states that all participants made their decisions before Dr. Lott's study was published"*	**Excuse: Excuse manipulation** *"The page states that all participants were told about Dr. Lott's study"*	**Control: Excuse manipulation** *(Blank)*

Donation decision
"Would you like to authorize a $1 donation to Fund the Wall?"

Post-treatment questions and perceived purpose

Debrief

Figure 4: Experiment 2: Donation rates by group
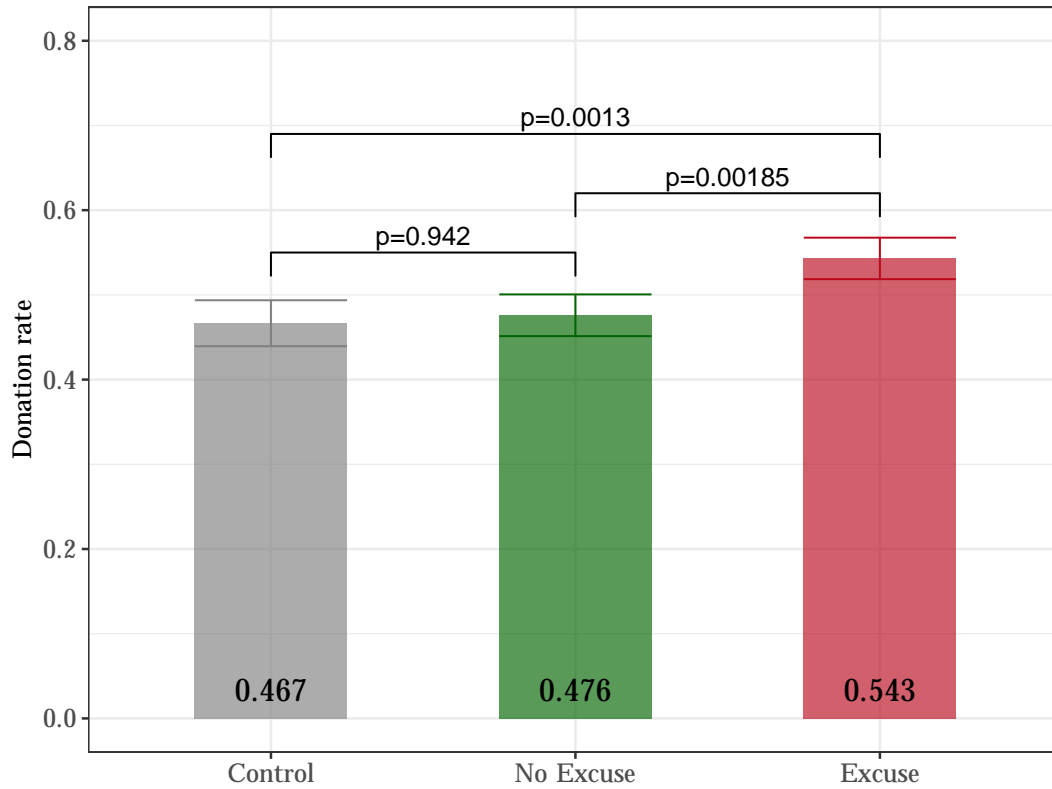
Notes: Figure 4 displays the results from Experiment 1, conducted in January 2020 with a sample of 3,728 Republicans and Independents. The figure displays donation rates to "Fund the Wall" across the Control group, the 'No Excuse' group, and the 'Excuse' group. The figure displays 95 percent confidence intervals as well as p-values for tests of equality of means across the conditions.

Tables

Table 1: Experiment 1: Inferred donation motives

	Dependent variable:					
	Inference about partner's donation motive					
	Used keyword			Predicted inference about score		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *Inference about intolerance*						
Excuse	−0.070***	−0.068***	−0.072***	−0.059***	−0.058***	−0.057***
	(0.012)	(0.012)	(0.009)	(0.017)	(0.017)	(0.013)
Constant	0.172***	0.291***	0.268***	0.168***	0.121	0.118
	(0.009)	(0.072)	(0.055)	(0.012)	(0.099)	(0.079)
Observations	3,047	3,047	5,065	3,045	3,045	5,061
R^2	0.010	0.029	0.025	0.004	0.023	0.021
Adjusted R^2	0.010	0.024	0.022	0.004	0.019	0.018
Panel B: *Inference about gullibility*						
Excuse	0.031***	0.032***	0.029***	0.109***	0.109***	0.101***
	(0.010)	(0.010)	(0.008)	(0.016)	(0.016)	(0.012)
Constant	0.069***	0.118**	0.113**	0.088***	0.019	0.027
	(0.007)	(0.058)	(0.045)	(0.011)	(0.091)	(0.072)
Observations	3,047	3,047	5,065	3,045	3,045	5,061
R^2	0.003	0.011	0.009	0.015	0.046	0.040
Adjusted R^2	0.003	0.006	0.006	0.015	0.041	0.037
Demographic controls	No	Yes	Yes	No	Yes	Yes
Include pilot data	No	No	Yes	No	No	Yes

Notes: The dependent variable in Columns 1-3 of Panel A is an indicator taking value 1 if the respondent uses a word relating to bias when describing why he or she thinks the matched respondent donated to Fund the Wall. The dependent variable in Columns 4-6 of Panel A is the predicted z-score of the (negative of the) Foreign Culture Tolerance Scale, where the prediction is based on the respondent's description of their matched respondent's motive. The dependent variables of Panel B are analogous, but instead consider words related to gullibility and the predicted score on the Gullibility Scale. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Table 2: Experiment 1: Inferred bias and gullibility scores

	Dependent variable:		
	Inference about partner's score		
	(1)	(2)	(3)
Panel A: *Bias (z-score)*			
Excuse	−0.134***	−0.133***	−0.152***
	(0.051)	(0.051)	(0.039)
Constant	0.058	−0.174	0.006
	(0.036)	(0.285)	(0.226)
Observations	1,524	1,524	2,532
R^2	0.004	0.038	0.037
Adjusted R^2	0.004	0.029	0.032
Panel B: *Gullibility (z-score)*			
Excuse	0.321***	0.310***	0.317***
	(0.050)	(0.050)	(0.039)
Constant	−0.155***	−0.100	−0.224
	(0.036)	(0.301)	(0.233)
Observations	1,523	1,523	2,533
R^2	0.026	0.065	0.059
Adjusted R^2	0.025	0.056	0.053
Demographic controls	No	Yes	Yes
Include pilot data	No	No	Yes

Notes: The dependent variable in Panel A is the negative of the z-score of the respondent's guess as to his or her matched respondent's score on the Foreign Culture Tolerance Scale, where we take the negative to interpret higher values as greater bias. The dependent variable in Panel B is the z-score of the respondent's guess as to his or her matched respondent's score on the Gullibility Scale. Both scales were originally scored between 0 and 100. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Table 3: Experiment 2: Main results

	Dependent variable:					
	Donated to Fund the Wall					
	(1)	(2)	(3)	(4)	(5)	(6)
Excuse	0.063***	0.065***	0.074***	0.066**	0.052**	0.071***
	(0.020)	(0.018)	(0.016)	(0.027)	(0.025)	(0.014)
Control	−0.001	−0.005	−0.001			0.017
	(0.020)	(0.018)	(0.017)			(0.016)
p-value (Excuse = Control)	0.0013	< 0.001	< 0.001			< 0.001
Demographic controls	No	Yes	Yes	No	Yes	Yes
Waves included	Main	Main	Main + Pilot	Replication	Replication	All
DV mean	0.488	0.488	0.497	0.497	0.497	0.498
DV std. dev.	0.500	0.500	0.500	0.500	0.500	0.500
Observations	3,751	3,751	4,457	1,373	1,373	5,913
R^2	0.004	0.187	0.197	0.004	0.130	0.171
Adjusted R^2	0.003	0.183	0.194	0.004	0.120	0.168

Notes: The dependent variable is an indicator taking value 1 if the respondent donated to Fund the Wall. Columns 1-2 report results estimated on the sample from the main experiment; Column 3 pools the sample from the main experiment with the sample from the pilot; Columns 4-5 consider only the sample from the replication experiment; and Column 6 pools all waves. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Table 4: Experiment 2: County heterogeneity

	Dependent variable:					
	Donated to Fund the Wall					
	(1)	(2)	(3)	(4)	(5)	(6)
Excuse	0.061***	0.063***	0.071***	0.041	0.026	0.064***
	(0.020)	(0.019)	(0.016)	(0.029)	(0.027)	(0.014)
Excuse × County Republican vote share	−0.030	−0.038**	−0.038**	−0.069**	−0.079***	−0.047***
	(0.021)	(0.019)	(0.017)	(0.027)	(0.026)	(0.014)
Control	−0.002	−0.005	−0.002			0.015
	(0.020)	(0.018)	(0.017)			(0.016)
Control × County Republican vote share	0.013	0.011	0.011			0.014
	(0.020)	(0.018)	(0.017)			(0.016)
County Republican vote share	0.052***	0.020	0.017	0.012	0.027	0.018*
	(0.015)	(0.013)	(0.012)	(0.019)	(0.019)	(0.010)
Demographic controls	No	Yes	Yes	No	Yes	Yes
Waves included	Main	Main	Main + Pilot	Replication	Replication	All
DV mean	0.488	0.488	0.497	0.497	0.497	0.498
DV std. dev.	0.500	0.500	0.500	0.500	0.500	0.500
Observations	3,631	3,631	4,315	1,215	1,215	5,608
R^2	0.013	0.192	0.203	0.009	0.133	0.176
Adjusted R^2	0.012	0.188	0.199	0.007	0.120	0.173

Notes: The dependent variable is an indicator taking value 1 if the respondent donated to Fund the Wall. The county Republican vote share is from the 2016 US Presidential election and is scaled to a standard normal distribution. Columns 1-2 include both Independents and Republicans, Columns 3-4 limit the sample to Republicans, and Columns 5-6 limit the sample to Independents. Columns 1, 3, and 5 report results estimated on the sample from the main experiment, while Columns 2, 4, and 6 pool the sample from the main experiment with the sample from the pilot. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, and a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Online Appendix

A Theory Proofs

A.1 Proof of Proposition 1

The tolerant sender $(i = 0)$ chooses to take the anti-immigrant action $(a = 1)$ if

$$v\mathbf{1}_{\{0=0\}} + \pi(0)b_0 + (1 - \pi(0))b_1 = u_0(0, \pi(0)) < u_0(1, \pi(1)) = v\mathbf{1}_{\{1=0\}} + \pi(1)b_0 + (1 - \pi(1))b_1$$

$$\Leftrightarrow v + \pi(0)b_0 + (1 - \pi(0))b_1 < \pi(1)b_0 + (1 - \pi(1))b_1$$

$$\Leftrightarrow v < (\pi(1) - \pi(0))(b_0 - b_1)$$

$$\Leftrightarrow \pi(1) - \pi(0) > \frac{v}{b_0 - b_1},$$

where the final inequality follows from the inequality $b_0 - b_1 > 0$. The intolerant sender $(i = 1)$ chooses to take the anti-immigrant action $(a = 1)$ if

$$v\mathbf{1}_{\{0=1\}} + \pi(0)b_0 + (1 - \pi(0))b_1 = u_0(0, \pi(0)) < u_0(1, \pi(1)) = v\mathbf{1}_{\{1=1\}} + \pi(1)b_0 + (1 - \pi(1))b_1$$

$$\Leftrightarrow \pi(0)b_0 + (1 - \pi(0))b_1 < v + \pi(1)b_0 + (1 - \pi(1))b_1$$

$$\Leftrightarrow -v < (\pi(1) - \pi(0))(b_0 - b_1)$$

$$\Leftrightarrow \pi(1) - \pi(0) > -\frac{v}{b_0 - b_1}.$$

A.2 Proof of Proposition 2

***No Excuse* game**   In the *No Excuse* game, the receiver believes that the sender has not seen the anti-immigrant information, so he expects the intolerant-persuadable sender to take the action and the tolerant-persuadable sender not to take the action. If both the tolerant-nonpersuadable and the intolerant-nonpersuadable senders do not take the action, Bayes' rule requires that $\pi^{NE}(1) = 0$ and $\pi^{NE}(0) = \frac{p}{p+q(1-p)}$. Letting S_i and G_i denote type-$i \in \{0, 1\}$ non-persuadable and persuadable senders, respectively, Bayes' rule gives:

$$\pi^{NE}(0) = \frac{\Pr(G_0, S_0)}{\Pr(S_0, S_1, G_0)} = \frac{(1-q)p + qp}{1 - (1-q)(1-p)} = \frac{p}{p + q - pq} = \frac{p}{p + q(1-p)}.$$

Because the tolerant-nonpersuadable sender does not take the action, the optimality condition for the intolerant-nonpersuadable sender, (2), yields the second inequality.

$$a_0^* = 0 \Leftrightarrow \pi^{NE}(1) - \pi^{NE}(0) = -\frac{p}{p + q(1-p)} \leq \frac{v}{b_0 - b_1},$$

$$a_1^* = 0 \Leftrightarrow -\frac{p}{p + q(1-p)} \leq -\frac{v}{b_0 - b_1}$$

$$\Leftrightarrow \frac{p}{p + q(1-p)} \geq \frac{v}{b_0 - b_1}$$

$$\Leftrightarrow v_1 \leq \frac{p(b_0 - b_1)}{p + q(1-p)}.$$

We now verify that no other pure-strategy equilibrium exists in the *No Excuse* condition. First, observe that if $a_0^* = 1$ then it must be that $a_1^* = 1$ from the optimality conditions. That is, we can rule out equilibria in which $a_0^* = 1$ and $a_1^* = 0$. It remains to rule out the following equilibria: (1) $a_0^* = 1$ and $a_1^* = 1$; and (2) $a_0^* = 0$ and $a_1^* = 1$.

(i) The receiver's posterior beliefs are:

$$\pi^{NE}(1) = \frac{\Pr(S_0)}{\Pr(S_0, S_1, G_1)} = \frac{qp}{1 - p(1-q)}, \quad \pi^{NE}(0) = 1$$

$$\Rightarrow \pi^{NE}(1) - \pi^{NE}(0) = \frac{qp}{1 - p(1-q)} - 1 = -\frac{1-p}{1 - p(1-q)} < 0.$$

This violates the optimality condition for S_0.

(ii) The receiver's posterior beliefs are:

$$\pi^{NE}(1) = 0, \ \pi^{NE}(0) = 1$$

$$\Rightarrow \pi^{NE}(1) - \pi^{NE}(0) = -1.$$

Thus, the optimality condition for S_0 is satisfied. For the optimality condition for S_1 to be satisfied, we need that

$$-1 > -\frac{v}{b_0 - b_1} \Leftrightarrow v > b_0 - b_1.$$

But this contradicts the hypothesis of Proposition 2, which implies that

$$v \leq \underbrace{\frac{p}{p + q(1-p)}}_{\in (0,1)} (b_0 - b_1) < b_0 - b_1 \Rightarrow v \leq b_0 - b_1.$$

***Excuse* game**   In the *Excuse* game, the receiver expects both types of persuadable senders to donate. Since we look for an equilibrium in which the tolerant-nonpersuadable sender does not donate and the intolerant-nonpersuadable sender donates, Bayes' rule requires $\pi^E(1) = \frac{p(1-q)}{1-pq}$ and $\pi^E(0) = 1$:

$$\pi^E(1) = \frac{\Pr(G_0)}{\Pr(G_0, G_1, S_1)} = \frac{(1-q)p}{(1-q)p + (1-q)(1-p) + q(1-p)} = \frac{p(1-q)}{1-qp} \in (0,1).$$

Because the tolerant-nonpersuadable sender does not donate, the intolerant-nonpersuadable sender's optimality condition yields the first inequality:

$$a_0^* = 0 \implies \pi^E(1) - \pi^E(0) = \frac{p(1-q)}{1-qp} - 1 \leq \frac{v}{b_0 - b_1}$$

$$\implies \frac{p(1-q) - 1 + qp}{1-qp} = -\frac{1-p}{1-qp} \leq \frac{v}{b_0 - b_1}$$

$$\implies -\frac{1-p}{1-qp} \leq 0 \leq \frac{v}{b_0 - b_1},$$

$$a_1^* = 1 \implies \frac{p(1-q)}{1-qp} - 1 = -\frac{1-p}{1-qp} > -\frac{v}{b_0 - b_1}$$

$$\implies \frac{1-p}{1-qp} < \frac{v}{b_0 - b_1}$$

$$\implies v > \frac{(1-p)(b_0 - b_1)}{1-qp}.$$

We appeal to the intuitive criterion to rule out equilibria in which both tolerant- and intolerant-nonpersuadable senders donate in equilibrium. In such an equilibrium, the receiver cannot use Bayes' rule if he observes that the sender does not donate. Given on-path belief, the first inequality implies that the intolerant-nonpersuadable sender cannot benefit by deviating to not donating, regardless of the receiver's belief. In particular, in this equilibrium, we have $\tilde{\pi}^E(1) = \Pr(G_0, S_0) = p$. The intolerant-nonpersuadable type's equilibrium payoff is

$$u_1\left(1, \pi^E(1)\right) = v + b\left(\tilde{\pi}^E(1)\right) = v + pb_0 + (1-p)b_1.$$

The best that she can do by deviating to $a = 0$ is:

$$\max_{\pi} \; u_1(0, \pi) = b(\pi) = b_0.$$

Hence, $a = 0$ is dominated if

$$v + pb_0 + (1 - p) b_1 > b_0 \Leftrightarrow (1 - p)(b_0 - b_1) < v.$$

This is satisfied by the hypothesis of Proposition 2, since

$$(1 - p)(b_0 - b_1) < \frac{(1 - p)(b_0 - b_1)}{1 - qp} < v.$$

Hence, by the intuitive criterion, the receiver must believe that any deviation from $a = 1$ is made by tolerant-nonpersuadable agents; i.e., $\tilde{\pi}^E(0) = 1$. In this case, we have

$$\tilde{\pi}^E(1) - \tilde{\pi}^E(0) = -(1 - p) < 0.$$

This violates the optimality condition for S_0, which rules out the possibility that both the tolerant- and intolerant-nonpersuadable senders donate in the (refined) equilibrium. Thus, the intuitive criterion requires the receiver to believe that the sender is intolerant if he observers the sender donating, which, in turn, implies that it is not optimal for the tolerant-nonpersuadable sender to donate.

We proceed to verify that other pure strategies cannot be part of any equilibrium. By the same argument in the *No Excuse* game, we can rule out the case in which $a_0^* = 1$ and $a_1^* = 0$. It remains to rule out the possibility that $a_0^* = 0$ and $a_1^* = 0$. In such an equilibrium,

$$\pi^E(0) = \frac{\Pr(S_0)}{\Pr(S_0, S_1)} = p, \ \pi^E(1) = \frac{\Pr(G_0)}{\Pr(G_0, G_1)} = p,$$

so that $\pi^E(1) - \pi^E(0) = 0$. But this violates the optimality condition for the intolerant-nonpersuadable sender, since $0 \nleq -\frac{v}{b_0 - b_1} < 0$.

The condition on q ensures that $0 < \frac{(1-p)(b_0-b_1)}{1-qp} \leq \frac{p(b_0-b_1)}{p+q(1-p)}$, i.e. that there exists some $v > 0$

that satisfies the set of inequalities in the statement of Proposition 2.

$$0 < \frac{p\,(b_0 - b_1)}{p + q\,(1 - p)} - \frac{(1 - p)\,(b_0 - b_1)}{1 - qp}$$

$$\implies \frac{1 - p}{1 - qp} < \frac{p}{p + q\,(1 - p)}$$

$$\implies (1 - p)\,(p + q\,(1 - p)) < p\,(1 - qp)$$

$$\implies p + q\,(1 - p) - p^2 - pq\,(1 - p) < p - qp^2$$

$$\implies q\,(1 - p) - p^2 - pq + qp^2 < -qp^2$$

$$\implies q - p^2 - 2pq + 2qp^2 < 0$$

$$\implies q\,(1 - 2p + 2p^2) < p^2$$

$$\implies q < \frac{p^2}{2p^2 - 2p + 1}.$$

B Appendix Figures and Tables

B.1 Experiment 1

Table B1: Experiment 1: Sample representativeness

	Experiment 1	Pew
Variables:	(1)	(2)
Age	41.60	46.67
Black	0.18	0.26
Asian	0.05	0.05
White	0.62	0.49
Hispanic	0.14	0.17
Male	0.47	0.39
Bachelors degree or higher	0.46	0.36
Observations	5065	4005

Notes: Table displays mean characteristics for the Experiment 1 sample and the 2018 Pew Research Center's American Trends Panel, Wave 39. Attriters dropped from sample.

Table B2: Experiment 1: Balance of covariates

	Overall		Excuse	No Excuse	p-value
	mean	std.dev.	mean	mean	(E=NE)
Variables:	(1)	(2)	(3)	(4)	(5)
Age	41.376	15.639	41.703	41.048	0.247
Black	0.182	0.386	0.186	0.179	0.612
Asian	0.045	0.208	0.049	0.042	0.386
White	0.710	0.454	0.703	0.716	0.455
Hispanic	0.140	0.347	0.136	0.144	0.561
Male	0.450	0.498	0.451	0.448	0.840
High school diploma	0.983	0.130	0.983	0.983	0.998
Bachelors degree	0.446	0.497	0.454	0.439	0.391

Notes: p-values based on robust standard errors reported. Attriters dropped from sample.

Table B3: Experiment 1: Relationship between perceived motive and scores

	Dependent variable:		
	Inference about partner's score		
	(1)	(2)	(3)
Panel A: *Bias (z-score)*			
Used bias word	0.477***	0.473***	0.500***
	(0.072)	(0.072)	(0.056)
Constant	−0.078***	−0.301	−0.149
	(0.027)	(0.281)	(0.222)
Observations	1,524	1,524	2,532
R^2	0.028	0.061	0.061
Adjusted R^2	0.027	0.052	0.056
Panel B: *Gullibility (z-score)*			
Used gullibility word	0.520***	0.484***	0.397***
	(0.089)	(0.088)	(0.070)
Constant	−0.039	−0.055	−0.107
	(0.026)	(0.302)	(0.234)
Observations	1,523	1,523	2,533
R^2	0.022	0.059	0.046
Adjusted R^2	0.021	0.051	0.041
Demographic controls	No	Yes	Yes
Include pilot data	No	No	Yes

Notes: The dependent variable in Panel A is the negative of the *z*-score of the respondent's guess as to his or her matched respondent's score on the Foreign Culture Tolerance Scale, where we take the negative to interpret higher values as greater bias. The dependent variable in Panel B is the *z*-score of the respondent's guess as to his or her matched respondent's score on the Gullibility Scale. Both scales were originally scored between 0 and 100. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Table B4: Experiment 1: Condition prediction confusion matrix

	Predicted Excuse	**Predicted Excuse**
True Excuse	212	185
True No Excuse	194	188

Overall accuracy: 0.5135

Notes: Each cell reports the number of individuals who were assigned to the condition (Excuse or No Excuse) in the corresponding row and who were classified by the Support Vector Machine as belonging to the condition in the corresponding column. The classifier was trained on a 75% sample of the data; the table reports prediction results on the test set of the remaining 25% of the data. Overall accuracy is calculated as the proportion of correct predictions.

Table B5: Experiment 2: Attrition

	Dependent variable:
	Attrited
	(1)
Age	0.001
	(0.002)
Age squared	0.00000
	(0.00003)
Black	−0.013
	(0.034)
Asian	0.008
	(0.045)
White	−0.038
	(0.030)
Hispanic	0.014
	(0.022)
Male	−0.030**
	(0.014)
High school	−0.023
	(0.049)
Some college, no degree	−0.035
	(0.048)
Associate degree	−0.047
	(0.051)

Continued on next page

	Attrited (1)
Bachelor degree	−0.067 (0.049)
Post-bachelor degree	−0.067 (0.051)
Strong Democrat	−0.021 (0.015)
Excuse × Age	−0.003 (0.003)
Excuse × Age squared	0.00002 (0.00003)
Excuse × Black	0.038 (0.047)
Excuse × Asian	−0.014 (0.062)
Excuse × White	0.049 (0.042)
Excuse × Hispanic	0.026 (0.031)
Excuse × Male	−0.004 (0.020)
Excuse × High school	0.017 (0.062)
Excuse × Some college, no degree	−0.0003 (0.060)
Excuse × Associate degree	−0.007 (0.066)
Excuse × Bachelor degree	0.018 (0.062)
Excuse × Post-bachelor degree	−0.005 (0.066)
Excuse × Strong Democrat	0.014 (0.022)
DV mean (no excuse)	0.093

Table B5 – *Continued from previous page*

	Attrited (1)
DV mean (excuse)	0.089
Observations	3,352
R^2	0.014
Adjusted R^2	0.007

Notes: The dependent variable is an indicator that takes value 1 if the respondent attrited post-randomization. Robust standard errors are reported.

B.2 Auxiliary Experiment

Table B6: Punishment of intolerant vs. gullible types

	Dependent variable:	
	Authorized \$1 bonus to partner	
	(1)	(2)
Gullible partner	0.112***	0.110***
	(0.039)	(0.039)
Constant	0.439***	0.166
	(0.028)	(0.275)
Demographic controls	No	Yes
Observations	646	646
R^2	0.012	0.054
Adjusted R^2	0.011	0.033

Notes: The dependent variable is an indicator taking value 1 if the respondent authorized a \$1 donation to their partner. The omitted group is subjects matched with an Intolerant (rather than Gullible) partner. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Figure B1: Punishment of intolerant vs. gullible types
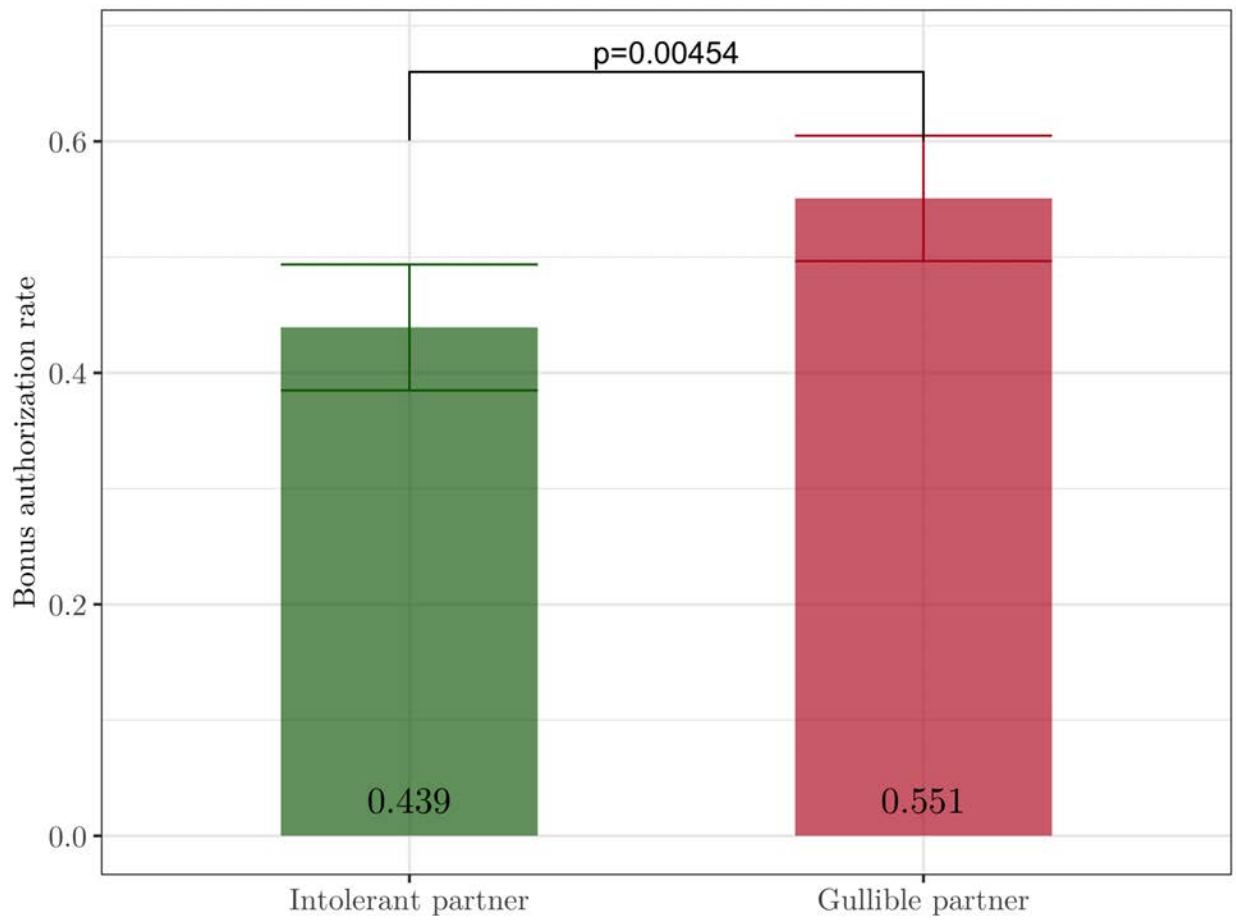
Notes: Figure B1 displays the results from the survey eliciting differential punishment of intolerant vs. gullible partners. The figure shows the fraction of respondents who chose to authorize a donation to their partner, separately by experimental condition, as well as 95 percent confidence intervals.

B.3 Experiment 2

Table B7: Experiment 2: Sample representativeness

|	Experiment 2	Pew
	(1)	(2)
Panel A: *Republican*		
Age	47.27	49.50
Black	0.03	0.02
Asian	0.03	0.03
White	0.84	0.84
Hispanic	0.09	0.08
Male	0.48	0.51
Bachelors degree or higher	0.38	0.29
Observations	1983	2879
Panel B: *Independent*		
Age	43.59	44.96
Black	0.11	0.08
Asian	0.05	0.04
White	0.69	0.70
Hispanic	0.12	0.13
Male	0.50	0.53
Bachelors degree or higher	0.37	0.34
Observations	2487	2622

Notes: Table displays mean respondent characteristics from the Experiment 2 sample and the 2018 Pew Research Center's American Trends Panel Wave, 39. Attriters dropped from sample.

Table B8: Experiment 2: Balance of covariates

	Overall		Excuse	No Excuse	Control	p-values		
	mean	std.dev.	mean	mean	mean	(E=NE)	(E=C)	(NE=C)
Variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	44.954	15.709	45.083	44.823	44.958	0.681	0.840	0.826
Black	0.076	0.266	0.070	0.088	0.072	0.100	0.858	0.135
Asian	0.043	0.203	0.041	0.042	0.046	0.877	0.567	0.677
White	0.821	0.383	0.825	0.815	0.823	0.498	0.858	0.609
Hispanic	0.110	0.313	0.113	0.108	0.110	0.691	0.829	0.852
Male	0.499	0.500	0.493	0.507	0.497	0.484	0.830	0.618
High school diploma	0.976	0.152	0.977	0.975	0.977	0.821	0.981	0.798
Bachelors degree	0.378	0.485	0.392	0.368	0.374	0.229	0.352	0.770
Republican	0.426	0.495	0.419	0.437	0.421	0.368	0.920	0.414

Notes: p-values based on robust standard errors reported. Attriters dropped from sample.

Figure B2: Experiment 2: Stability of estimated treatment effect
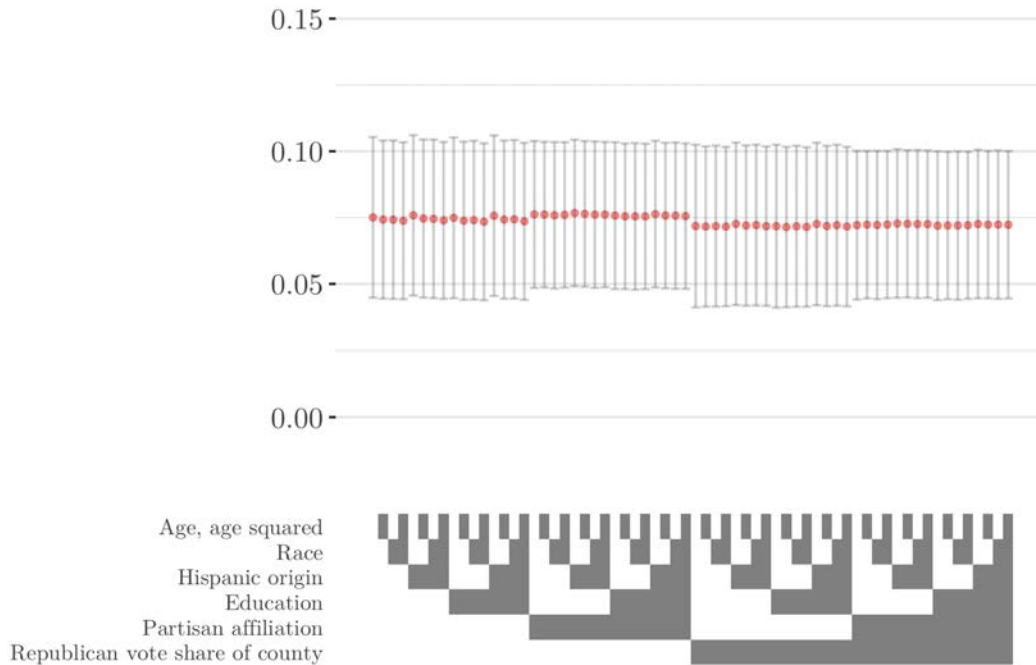
Notes: Figure B2 displays the estimated treatment effects of the *Excuse* condition (relative to the *No Excuse* condition) on donation rates to Fund the Wall under every possible set of demographic controls. 95% confidence intervals are reported.

Table B9: Experiment 2: Party heterogeneity

	Dependent variable:			
	Donated to Fund the Wall			
	Republicans		Independents	
	(1)	(2)	(3)	(4)
Excuse	0.067**	0.078***	0.068***	0.074***
	(0.028)	(0.024)	(0.024)	(0.023)
Control	−0.009	−0.010	0.003	0.012
	(0.028)	(0.026)	(0.024)	(0.024)
Observations	1,593	1,973	2,158	2,484
Demographic controls	Yes	Yes	Yes	Yes
p-value (Excuse = Control)	0.0076	< 0.001	0.0055	0.0082
Include pilot data	No	Yes	No	Yes
DV mean	0.669	0.673	0.354	0.357
DV std. dev.	0.471	0.469	0.478	0.469
R^2	0.070	0.050	0.132	0.048
Adjusted R^2	0.061	0.043	0.126	0.042

Notes: The dependent variable is an indicator taking value 1 if the respondent donated to Fund the Wall. Columns 1-2 limit the sample to Republicans, while Columns 3-4 limit the sample to Independents. Columns 1-2 report results estimated on the sample from the main experiment; Column 3 pools the sample from the main experiment with the sample from the pilot; Columns 4-5 consider only the sample from the replication experiment; and Column 6 pools all waves. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Table B10: Experiment 2: Perceived purpose of study

	Excuse (1)	Immigration attitudes (2)	Public image (3)	Information (4)	Persuasion (5)	Biased (6)
			Dependent variable:			
Excuse	−0.005 (0.003)	0.009 (0.015)	0.020** (0.010)	0.012 (0.015)	−0.013 (0.011)	−0.0001 (0.013)
Control	−0.003 (0.003)	0.129*** (0.015)	0.036*** (0.010)	−0.015 (0.016)	−0.081*** (0.012)	−0.037*** (0.014)
p-value (Excuse = Control)	0.63	< 0.001	0.12	0.093	< 0.001	0.0082
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Include pilot data	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.007	0.226	0.082	0.239	0.121	0.175
DV std. dev.	0.084	0.419	0.275	0.427	0.326	0.380
Observations	4,537	4,537	4,537	4,536	4,537	4,535
R^2	0.004	0.027	0.018	0.011	0.023	0.009
Adjusted R^2	0.001	0.024	0.014	0.007	0.019	0.006

Notes: The dependent variable in each column is an indicator taking value 1 if the respondent's perceived purpose of the study was coded to fall into the corresponding category. "Excuse" takes value 1 if the respondent correctly inferred the study was about whether knowing that others will know one had an "excuse" for donating would affect the donation decision. "Immigration attitudes" takes value 1 if the respondent stated the study was about attitudes toward immigration. "Public image" takes value 1 if the respondent stated the study was about whether knowing one's decision will be observable to others would affect the donation decision. "Information" takes value 1 if the respondent stated the study was about disseminating information about immigration. "Persuasion" takes value 1 if the respondent stated the researchers were attempting to persuade them either to donate or not to donate. "Bias" takes value 1 if the respondent stated the researchers were biased. "Other" takes value 1 if the respondent stated a purpose that did not fall into any of the above categories. Categories other than "Other" are not mutually exclusive. All specifications pool the main experiment and the pilot and contol for demographics and partisan affiliation. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, and a set of education indicators. Partisan affiliation controls include dummies for strong Republican, weak Republican, Republican-leaning Independent, and Democrat-leaning Independent. Robust standard errors are reported.

Table B12: Experiment 2: Attrition

	Dependent variable:
	Attrited
	(1)
Age	−0.002
	(0.003)
Age squared	0.0001*
	(0.00003)
Black	−0.012
	(0.056)
Asian	0.062
	(0.064)
White	0.034
	(0.046)
Hispanic	−0.032
	(0.034)
Male	−0.067***
	(0.020)
High school	−0.024
	(0.061)
Some college, no degree	−0.042
	(0.060)
Associate degree	−0.051
	(0.063)
Bachelor degree	−0.023
	(0.060)
Post-bachelor degree	−0.077
	(0.064)
Rep-leaning Ind	0.009
	(0.026)

Continued on next page

Table B12 – *Continued from previous page*

	Attrited
	(1)
Weak Rep	−0.030
	(0.032)
Strong Rep	−0.039
	(0.026)
Excuse × Age	−0.003
	(0.004)
Excuse × Age squared	0.00003
	(0.00004)
Excuse × Black	0.078
	(0.075)
Excuse × Asian	−0.049
	(0.089)
Excuse × White	0.0003
	(0.060)
Excuse × Hispanic	0.028
	(0.047)
Excuse × Male	0.007
	(0.028)
Excuse × High school	0.071
	(0.081)
Excuse × Some college, no degree	0.039
	(0.080)
Excuse × Associate degree	0.081
	(0.085)
Excuse × Bachelor degree	0.081
	(0.081)
Excuse × Post-bachelor degree	0.096
	(0.087)

Continued on next page

Table B12 – *Continued from previous page*

| | Attrited |
	(1)
Excuse × Rep-leaning Ind	−0.014
	(0.036)
Excuse × Weak Rep	0.045
	(0.045)
Excuse × Strong Rep	0.008
	(0.036)
DV mean (no excuse)	0.166
DV mean (excuse)	0.177
Observations	3,014
R^2	0.035
Adjusted R^2	0.025

Notes: The dependent variable is an indicator that takes value 1 if the respondent attrited post-randomization. The sample is limited to respondents in the Excuse and No Excuse condition. Robust standard errors are reported.

Table B11: Experiment 2: Condition prediction confusion matrices

Panel A: *Excuse vs. No Excuse*

	Predicted Excuse	Predicted Excuse
True Excuse	213	244
True No Excuse	210	210

Overall accuracy: 0.4823

Panel B: *Control vs. No Excuse*

	Predicted Excuse	Predicted Excuse
True Control	197	180
True No Excuse	136	283

Overall accuracy: 0.6030

Panel C: *Control vs. Excuse*

	Predicted Excuse	Predicted Excuse
True Control	188	159
True Excuse	136	315

Overall accuracy: 0.6303

Notes: Each cell reports the number of individuals who were assigned to the condition in the corresponding row and who were classified by the Support Vector Machine as belonging to the condition in the corresponding column. Each panel limits the data to the corresponding two conditions. The classifiers were trained on a 75% sample of the limited dataset; the table reports prediction results on the test set of the remaining 25% of the limited dataset. Overall accuracy is calculated as the proportion of correct predictions.

C Additional Survey

In May 2020, we conducted a survey among a broadly representative sample of 1,121 Republicans in partnership with Luc.id.[55]

Design At the time of the experiment, the United States and the world were grappling with the COVID-19 pandemic, and many nations, including the US, had implemented restrictions on international travel in order to limit the spread of the virus. The Trump administration also cited the COVID-19 pandemic and the economic crisis it precipitated as a justification for severely limiting immigration, limiting the issuance of green cards and eliminating the STEM OPT program, which allowed graduates of US undergraduate institutions to remain in the country for one to three years to work in a related field.[56] A number of commentators argued that people were using COVID-19 as an excuse for xenophobic behavior against Asian-Americans; moreover, President Trump also claimed that the pandemic exacerbated the crisis on the US–Mexico border, renewing calls for the construction of a border wall.[57]

Motivated by these justifications, we inform participants that we are interested in their views on whether the United States should implement a *permanent* ban on immigration from Mexico. We tell participants that we will post their individual donation decision on our public study website, showing them a screenshot of an example of the website displaying respondents' support for the permanent ban on Mexican immigration. Participants are randomized into one of two conditions. Those in the *Excuse* condition are asked whether they agree with the statement that "I support a permanent ban on Mexican immigration to protect the US from contagious diseases, such as the coronavirus," and are informed that if they indicate that they agree with the statement, this is the wording that will appear on the study website. Those in the *No Excuse* condition are asked whether they agree with the statement that "I support a permanent ban on Mexican immigration," and, again, are informed that if they agree, this is the wording that will appear on the website. Thus, the key difference between the two conditions is whether a public justification is attached to

---

[55]All survey instruments are available in Appendix G. Appendix Table C1 presents summary statistics and compares our sample to the Pew American Trends Panel, confirming that our sample is indeed broadly representative. Appendix Table C2 confirms that demographic characteristics of respondents are balanced across treatment conditions.

[56]See, for example, "Next Trump Immigration Target: OPT For International Students." *Forbes*, May 4, 2020.

[57]See, for example, "Another Danger of COVID-19: Discrimination". *UNICEF USA*, April 10, 2020

respondents' support for the statement.[58]

Results The findings, displayed in Figure C1 and Table C3, are striking: 32 percent of respondents publicly support a permanent ban on Mexican immigration when they do not have an excuse, while 51 percent of respondents support a permanent ban on Mexican immigration when provided with the excuse — a 59 percent, statistically significant increase ($p < 0.001$). These findings underscore the quantitative importance of simple excuses as catalysts for xenophobic expression.

Table C1: Motivating survey: Sample representativeness

	Survey	Pew
Variables:	(1)	(2)
Age	46.68	49.50
Black	0.02	0.02
Asian	0.04	0.03
White	0.85	0.84
Hispanic	0.08	0.08
Male	0.40	0.51
Some college, no degree	0.23	0.24
Bachelors degree	0.30	0.18
Observations	1121	2879

Notes: Mean of respondent characteristics in the motivating study and the 2018 Pew Research Center's American Trends Panel Wave 39. Attriters dropped from sample.

---

[58]To rule out that the effect in the *Excuse* condition is driven by an increased salience of the disease protection rationale, we ask all respondents in the *No Excuse* condition before asking the main question: "Do you worry that Mexican immigration could make the country more vulnerable to contagious diseases, such as the coronavirus?"

Table C2: Motivating survey: Balance of covariates

| Variables: | Overall | | Excuse | No Excuse | p-value |
| | mean | std.dev. | mean | mean | (E=NE) |
	(1)	(2)	(3)	(4)	(5)
Age	46.682	16.451	46.716	46.649	0.946
Black	0.023	0.151	0.023	0.023	0.960
Asian	0.039	0.194	0.046	0.032	0.245
White	0.901	0.299	0.901	0.901	0.993
Hispanic	0.084	0.277	0.088	0.079	0.585
Male	0.402	0.491	0.396	0.409	0.651
High school diploma	0.991	0.094	0.989	0.993	0.546
Bachelors degree	0.301	0.459	0.307	0.294	0.617

Notes: p-values based on robust standard errors reported. Attriters dropped from sample.

Table C3: Motivating survey: results

| | *Dependent variable:* | | |
| | Publicly supports permanent ban | | |
	(1)	(2)	(3)
Excuse	0.195***	0.197***	0.195***
	(0.029)	(0.029)	(0.028)
Demographic controls	No	Yes	Yes
Partisan affiliation controls	No	No	Yes
DV mean	0.416	0.416	0.416
DV std. dev.	0.493	0.493	0.493
Observations	1,121	1,121	1,121
R^2	0.039	0.080	0.143
Adjusted R^2	0.038	0.065	0.129

Notes: The dependent variable is an indicator taking value 1 if the respondent agreed that immigration from Mexico should be permanently banned. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, a set of education indicators, and a set of partisan affiliation indicators. Robust standard errors are reported.

Figure C1: Covid Excuse
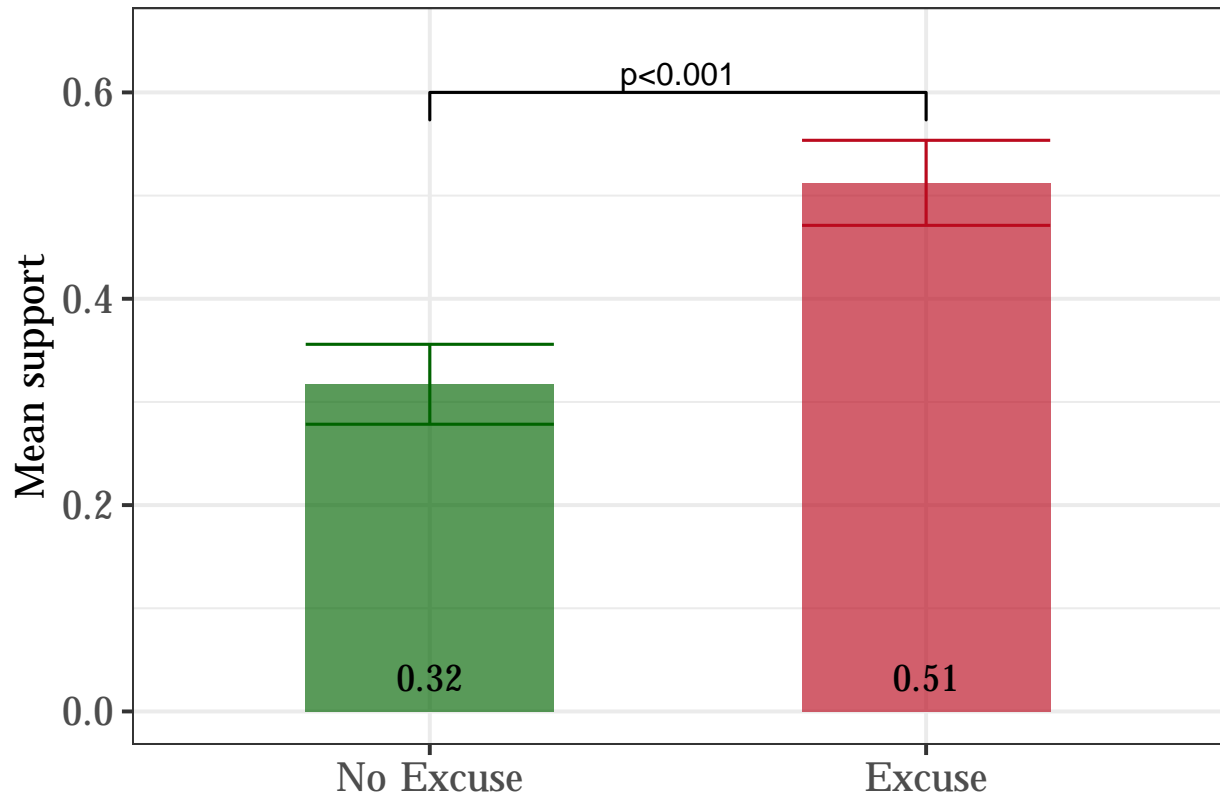
Notes: Figure C1 displays the results from the motivating survey experiment conducted in May 2020 with a sample of 1,121 Republicans. The figure shows the fraction of respondents supporting a permanent ban of Mexican immigration in the *Excuse* condition and in the *No Excuse* condition as well as 95 percent confidence intervals.

Supplementary Materials:
Experimental Instructions

Not intended for publication

D Survey instruments: Experiment 1

D.1 Consent and pre-treatment questions

We are a group of **nonpartisan researchers** interested in compiling an accurate and unbiased report about political and social attitudes in the US. This survey will take **approximately 5 minutes**.

Before proceeding, please note that it is important for our survey that you read all questions carefully and answer as accurately as possible. If your answers indicate that you did not carefully read questions, we will be forced to terminate your survey and will be unable to pay you.

Consent for Participation in a Research Study
Study Title: Political and Social Attitudes
Principal Investigator: Leonardo Bursztyn
IRB Study Number: IRB19-1320

DESCRIPTION: We are doing a research study about political and social attitudes in the United States. The research project will consist of reading information and answering a few short questions. Participation should take approximately five minutes.

RISKS and BENEFITS: The risks to your participation in this online study are those associated with basic computer tasks, including boredom, fatigue, mild stress, or breach of confidentiality. The only benefit to you, other than survey compensation, is the learning experience from participating in a research study. The benefit to society is the contribution to scientific knowledge.

COMPENSATION: Upon completion of the study, you will receive compensation in the amount you have agreed to with the platform through which you entered this survey. Partially-completed survey responses will not be compensated.

CONFIDENTIALITY: All data will be stored on password-protected servers and hard drives. We do not ask for any identifying information.

Any reports and presentations about the findings from this study will not include any identifying information. We may share the data we collect in this study with other researchers doing future studies – if we share your data, we will not include information that could identify you.

SUBJECT'S RIGHTS: Your participation is voluntary. You may stop participating at any time by closing the browser window or the program to withdraw from the study.

For additional questions about this research, you may contact:
Leonardo Bursztyn, bursztyn.research@gmail.com

For questions about your rights as a research participant, you may contact:
The Social & Behavioral Sciences Institutional Review Board, University of Chicago
Phone: (773) 834-7835
E-mail: sbs-irb@uchicago.edu

Please indicate, in the box below, that you are at least 18 years old, have read and understand this consent form, and you agree to participate in this online research study.

I have read and understood the above and want to participate in this study.

○ Yes

○ No

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. **To show that you read our questions carefully, please choose both "Extremely interested" and "Not at all interested" as your answer in the next question.** How interested are you in sports?

☐ Extremely interested

☐ Very interested

☐ A little bit interested

☐ Almost not interested

☐ Not at all interested

›

Are you Spanish, Hispanic, or Latino or none of these?

○ Yes

○ None of these

What is your year of birth?

[▾]

What is your sex?

○ Male

○ Female

In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?

○ Republican

○ Democrat

○ Independent

››

Do you strongly support or weakly support the Democratic Party?

○ Strongly support

○ Weakly support

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)

Which of the following best describes your race or ethnicity?

○ African American/Black

○ Asian/Asian American

○ Caucasian/White

○ Native American, Inuit or Aleut

○ Native Hawaiian/Pacific Islander

○ Other

>>

Do Illegal Immigrants Commit More Crime? Evidence from Dr. Lott's Study

Dr. John R. Lott, an economist formerly employed at top institutions such as Yale University and the University of Chicago, carried out a study on the relationship between illegal immigration and crime using new high-quality data. **The study has not yet been published in an academic journal, but we obtained an early version and summarize the results below.**

Dr. Lott's study uses **detailed data spanning over two decades from the Arizona state prison systems**. He finds that **"undocumented immigrants are at least 142% more likely** to be convicted of a crime than other Arizonans. They also tend to commit **more serious crimes**," such as murder and rape.

Moreover, "young undocumented immigrants commit crime at twice the rate of young U.S. citizens." He calculates that "if undocumented immigrants committed crime nationally as they do in Arizona, in 2016 they would have been responsible for over **1,000 more murders, 5,200 rapes, 8,900 robberies, 25,300 aggravated assaults, and 26,900 burglaries**. The vast majority of the illegal immigrants Lott studies are from Mexico.

However, a number of sources (including a researcher affiliated with the Cato Institute, a non-partisan think tank) have recently challenged some of the study's methods, claiming that errors in analysis invalidate its results. Dr. Lott has responded to this critique with a defense of the study's methods, but the issue remains unresolved.

›››

D.2   *No Excuse* condition

We conducted a survey about political and social attitudes in the United States earlier this year. You have been matched with one of the respondents from that survey.

We gave your matched respondent the opportunity to authorize a $1 donation to **Fund the Wall**, a nonprofit organization that seeks to reduce illegal immigration into the United States by helping to fund and construct the US-Mexico border wall. Your matched respondent was told that their donation decision would be posted on our website. The decision on whether to authorize the donation did not have any financial consequences for your matched respondent.

Some respondents were assigned a longer version of the survey and learned about Dr. Lott's study before they decided whether or not to donate. Other respondents were assigned a shorter version of the study and **were not informed** about Dr. Lott's study before they decided whether or not to donate.

Information about your matched respondent

- Your matched respondent **was not informed about Dr. Lott's study**, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent decided to authorize the $1 donation to Fund the Wall

››

Why do you think your matched respondent chose to donate to Fund the Wall?

[ ]

Reminder: Information about your matched respondent

- Your matched respondent **was not informed about Dr. Lott's study**, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent decided to authorize the $1 donation to Fund the Wall

››

After your matched respondent made their donation decision, they completed the **The Gullibility Scale**, a short questionnaire which measures **how easily people are manipulated by evidence from untrustworthy sources**.

On the next page, we will ask you to guess how your matched respondent scored on this scale. If you guess the correct option, you will be entered into a lottery for a $50 Amazon gift card.

››

The Gullibility Scale

We administered **The Gullibility Scale**, a short questionnaire which measures **how easily people are manipulated by evidence from untrustworthy sources**, to your matched respondent.

The test is scored from 0 to 100, where 0 means "least gullible" and 100 means "most gullible". Thus, a higher score indicates that your matched respondent is more gullible.

Reminder: Information about your matched respondent

- Your matched respondent **was not** informed about Dr. Lott's study, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent decided to authorize the $1 donation to Fund the Wall

If you had to guess, how do you think your **matched respondent** scored on **The Gullibility Scale**?

○ Score between 0 and 10 (**Not at all gullible**)

○ Score between 10 and 20

○ Score between 20 and 30

○ Score between 30 and 40

○ Score between 40 and 50

○ Score between 50 and 60

○ Score between 60 and 70

○ Score between 70 and 80

○ Score between 80 and 90

○ Score between 90 and 100 (**Extremely gullible**)

>>

After your matched respondent made their donation decision, they completed the **Foreign Culture Tolerance Scale**, a short questionnaire which measures **tolerance toward foreign values and traditions**.

On the next page, we will ask you to guess how your matched respondent scored on this scale. If you guess the correct option, you will be entered into a lottery for a $50 Amazon gift card.

>>

The Foreign Culture Tolerance Scale

We administered the **Foreign Culture Tolerance Scale**, a short questionnaire which measures tolerance toward **foreign values and traditions**, to your matched respondent.

The test is scored from 0 to 100, where 0 means "least tolerant" and 100 means "most tolerant". Thus, a **higher score indicates that your matched respondent is more tolerant toward foreign values and traditions.**

Reminder: Information about your matched respondent

- Your matched respondent **was not** informed about Dr. Lott's study, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent decided to authorize the $1 donation to Fund the Wall

If you had to guess, how do you think your **matched respondent** scored on the **Foreign Culture Tolerance Scale**?

○ Score between 0 and 10 (**Not at all tolerant**)

○ Score between 10 and 20

○ Score between 20 and 30

○ Score between 30 and 40

○ Score between 40 and 50

○ Score between 50 and 60

○ Score between 60 and 70

○ Score between 70 and 80

○ Score between 80 and 90

○ Score between 90 and 100 (**Extremely tolerant**)

››

77

D.3 *Excuse* condition

We conducted a survey about political and social attitudes in the United States earlier this year. You have been matched with one of the respondents from that survey.

We gave your matched respondent the opportunity to authorize a $1 donation to **Fund the Wall**, a nonprofit organization that seeks to reduce illegal immigration into the United States by helping to fund and construct the US–Mexico border wall. Your matched respondent was told that their donation decision would be posted on our website. The decision on whether to authorize the donation did not have any financial consequences for your matched respondent.

Some respondents were assigned a longer version of the survey and learned about Dr. Lott's study before they decided whether or not to donate. Other respondents were assigned a shorter version of the study and **were not informed** about Dr. Lott's study before they decided whether or not to donate.

Information about your matched respondent

- Your matched respondent **was informed about Dr. Lott's study**, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent then decided to authorize the $1 donation to Fund the Wall

››

78

Why do you think your matched respondent chose to donate to Fund the Wall?

[ ]

Reminder: Information about your matched respondent

- Your matched respondent **was informed about Dr. Lott's study**, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent then decided to authorize the $1 donation to Fund the Wall

››

After your matched respondent made their donation decision, they completed the **The Gullibility Scale**, a short questionnaire which measures **how easily people are manipulated by evidence from untrustworthy sources**.

On the next page, we will ask you to guess how your matched respondent scored on this scale. If you guess the correct option, you will be entered into a lottery for a $50 Amazon gift card.

››

The Gullibility Scale

We administered **The Gullibility Scale**, a short questionnaire which measures **how easily people are manipulated by evidence from untrustworthy sources**, to your matched respondent.

The test is scored from 0 to 100, where 0 means "least gullible" and 100 means "most gullible". Thus, a higher score indicates that your matched respondent is more gullible.

Reminder: Information about your matched respondent

- Your matched respondent **was informed about Dr. Lott's study**, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent then decided to authorize the $1 donation to Fund the Wall

If you had to guess, how do you think your **matched respondent** scored on **The Gullibility Scale**?

○ Score between 0 and 10 (**Not at all gullible**)

○ Score between 10 and 20

○ Score between 20 and 30

○ Score between 30 and 40

○ Score between 40 and 50

○ Score between 50 and 60

○ Score between 60 and 70

○ Score between 70 and 80

○ Score between 80 and 90

○ Score between 90 and 100 (**Extremely gullible**)

››

After your matched respondent made their donation decision, they completed the **Foreign Culture Tolerance Scale**, a short questionnaire which measures **tolerance toward foreign values and traditions**.

On the next page, we will ask you to guess how your matched respondent scored on this scale. If you guess the correct option, you will be entered into a lottery for a $50 Amazon gift card.

The Foreign Culture Tolerance Scale

We administered the **Foreign Culture Tolerance Scale**, a short questionnaire which measures tolerance toward **foreign values and traditions**, to your matched respondent.

The test is scored from 0 to 100, where 0 means "least tolerant" and 100 means "most tolerant". Thus, a **higher score indicates that your matched respondent is more tolerant toward foreign values and traditions.**

Reminder: Information about your matched respondent

- Your matched respondent **was informed about Dr. Lott's study**, which finds that illegal immigrants commit more crimes than US citizens
- Your matched respondent then decided to authorize the $1 donation to Fund the Wall

If you had to guess, how do you think your **matched respondent** scored on the **Foreign Culture Tolerance Scale**?

○ Score between 0 and 10 (**Not at all tolerant**)

○ Score between 10 and 20

○ Score between 20 and 30

○ Score between 30 and 40

○ Score between 40 and 50

○ Score between 50 and 60

○ Score between 60 and 70

○ Score between 70 and 80

○ Score between 80 and 90

○ Score between 90 and 100 (**Extremely tolerant**)

>>

82

D.4   Post-treatment questions and debrief

Before today, had you taken any previous online surveys that discussed Dr. Lott's study about the crime rates of illegal immigrants?

○ Yes

○ No

〉〉

If you had to guess, what would you say was the purpose of this study?

〉〉

If you have any feedback on our survey, please leave it below.

〉〉

Thanks for completing all our questions!

Study Title: Political and Social Attitudes
Principal Investigator: Leonardo Bursztyn
IRB Study Number: IRB19-1320

Debrief about crime rates among immigrants in the survey

We earlier provided you with truthful information about Dr Lott's study. As we mentioned earlier in the survey, Dr. Lott's study has been challenged by a number of sources for inaccuracies in data analysis. While his methods have not been entirely debunked, there remains a great deal of controversy. Due to these problems, it is unclear whether Dr. Lott's results will be published in a reputable academic journal. We did not expand upon this controversy during the study, but if you wish to read more, we suggest this analysis by a researcher at the Cato Institute, a nonpartisan think-tank.

Immigration and crime refers to perceived or actual relationships between crime and immigration. The academic literature provides mixed findings for the relationship between immigration and crime worldwide, but finds for the United States that immigration either has no impact on the crime rate or that it reduces the crime rate. A meta-analysis of 51 studies from 1994–2014 on the relationship between immigration and crime in the United States found that overall immigration reduces crime, but the relationship is very weak. Research suggests that people tend to overestimate the relationship between immigration and criminality,and that the media tends to erroneously depict immigrants as particularly crime-prone.

The relevant meta-analysis we are referring to is the following article.
Ousey, Graham C.; Kubrin, Charis E. (2018). "Immigration and Crime: Assessing a Contentious Issue". *Annual Review of Criminology*. **1** (1): null.

To get more information on this meta-analysis, clock on the link below:
doi:10.1146/annurev-criminol-032317-092026.

Contacts & Questions:
If you have questions or concerns about the study, you can contact the researchers at bursztyn.research@gmail.com.

Final Report: If you would like to receive a report of this study (or a summary of the findings) when it is completed, contact the researcher at the email address or phone number above.

›

E   Survey instruments: Experiment 2A

Moreover, I consent to researchers at Harvard University and the University of Chicago accessing the following information from my study provider: first and last name, city, and operating system (Windows, Mac, iOS, Android, or other). **We will not have access to any other personally-identifying information, and the information will be used for study purposes only.**

○ Yes

○ No

››

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. **To show that you read our questions carefully, please choose both "Extremely interested" and "Not at all interested" as your answer in the next question.** How interested are you in sports?

☐ Extremely interested

☐ Very interested

☐ A little bit interested

☐ Almost not interested

☐ Not at all interested

Are you Spanish, Hispanic, or Latino or none of these?

○ Yes

○ None of these

What is your year of birth?

[:]

What is your sex?

○ Male

○ Female

In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?

- ○ Republican
- ○ Democrat
- ○ Independent

››

Do you strongly support or weakly support the Republican Party?

- ○ Strongly support
- ○ Weakly support

››

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)

Which of the following best describes your race or ethnicity?

- ○ African American/Black
- ○ Asian/Asian American
- ○ Caucasian/White
- ○ Native American, Inuit or Aleut
- ○ Native Hawaiian/Pacific Islander
- ○ Other

››

Please read the following information carefully, as it is important for the rest of the survey.

››

Studying the crime rates of illegal immigrants is difficult. As the independent fact-checking organization PolitiFact reported in 2016, "The challenge in finding concrete numbers is due to a shortfall of data. There is no national database or study tracking how many people have been killed by undocumented immigrants or the nationality of the victims."

Therefore, most previous research has used flawed data.

Information about Lott Study: *Excuse* and *No Excuse* condition

Do Illegal Immigrants Commit More Crime? Evidence from Dr. Lott's Study

Dr. John R. Lott, an economist formerly employed at top institutions such as Yale University and the University of Chicago, carried out a study on the relationship between illegal immigration and crime using new high-quality data. **The study has not yet been published in an academic journal, but we obtained an early version and summarize the results below.**

Dr. Lott's study uses **detailed data spanning over two decades from the Arizona state prison systems**. He finds that "undocumented immigrants are at least 142% more likely to be convicted of a crime than other Arizonans. They also tend to commit **more serious crimes**," such as murder and rape.

Moreover, "young undocumented immigrants commit crime at twice the rate of young U.S. citizens." He calculates that "if undocumented immigrants committed crime nationally as they do in Arizona, in 2016 they would have been responsible for over **1,000 more murders, 5,200 rapes, 8,900 robberies, 25,300 aggravated assaults, and 26,900 burglaries**. The vast majority of the illegal immigrants Lott studies are from Mexico.

However, a number of sources (including a researcher affiliated with the Cato Institute, a non-partisan think tank) have recently challenged some of the study's methods, claiming that errors in analysis invalidate its results. Dr. Lott has responded to this critique with a defense of the study's methods, but the issue remains unresolved.

Do you think that Dr. Lott's study will be widely discussed in the media when it is published in an academic journal?

○ Yes

○ No

87

Reconsent

Before proceeding, please re-confirm your consent to the following:

I consent to researchers at Harvard University and University of Chicago accessing the following information from my study provider: first and last name, city, and operating system (Windows, Mac, iOS, Android, or other). We will not have access to any other personally-identifying information, and the information will be used for study purposes only.

○ Yes, I consent and would like to proceed with the survey.

○ No, I do not consent and would like to terminate the survey now.

Description of donation decision

We will now **randomly** select one of two organizations, and you will have the opportunity to authorize a $1 donation to this organization. We will make the donation on your behalf, so **it will not be deducted from your payment**.

- One organization seeks to reduce illegal immigration into the United States by raising money for and helping construct the US-Mexico border wall.
- The other organization seeks to aid families detained at the border by providing them with legal assistance and counsel.

On the next screen, you will be shown which organization has been selected.

Donation decision

The organization randomly selected for you is **Fund the Wall**.

Fund the Wall is a nonprofit organization that seeks to reduce illegal immigration into the United States by helping to fund and construct the US-Mexico border wall. If you wish, you can choose to authorize a $1 donation to Fund the Wall.

We will make the donation on your behalf, so **it will not be deducted from your payment**.

88

Website *Excuse* condition and *No Excuse* condition

Publicizing the results of our study

As researchers, we believe it is important to communicate our findings about political and social attitudes in Mannheim to the public.

Therefore, once Dr. Lott's study is published in a reputable academic journal, we will post the results from this survey, **including your individual donation decision and the donation decisions of all of the other respondents to this survey who consented on the previous page**, on our website.

We will then work with major news organizations in Mannheim with both a liberal and conservative viewership to publicize our website through newspaper and website articles, and we will also promote our website via Facebook ads to Mannheim residents.

What website visitors will learn

As you can see on the screenshot below, website visitors learn about Dr. Lott's study.

Do Illegal Immigrants Commit More Crime? Evidence from Dr. Lott's Study

Dr. John R. Lott, an economist formerly employed at top institutions such as Yale University and the University of Chicago, has posted a study on illegal immigration and crime. Dr. Lott's study **uses detailed data spanning over two decades** from the Arizona state prison systems. He finds that "**undocumented immigrants are at least 142% more likely** to be convicted of a crime than other Arizonans. They also tend to commit **more serious crimes**," such as murder and rape. Moreover, "young undocumented immigrants commit crime at **twice the rate** of young U.S. citizens."

He calculates that "if undocumented immigrants committed crime nationally as they do in Arizona, in 2016 they would have been responsible for over **1,000 more murders, 5,200 rapes, 8,900 robberies, 25,300 aggravated assaults, and 26,900 burglaries**." The vast majority of the illegal immigrants Lott studies are from Mexico. Given that there have been relatively few academic studies using high-quality data, Dr. Lott's study is among the **first of its kind**. However, a number of sources (including the Cato Institute, a non-partisan think tank) have recently challenged some of the study's methods, claiming that errors in analysis invalidate its results. Dr. Lott has responded to this critique with a defense of the study's methods, but the issue remains unresolved.

What do you think website visitors will conclude about Dr. Lott's study after visiting the web pagepage?

- ○ That Dr. Lott finds that illegal immigrants are less likely to commit crimes than US citizens.

- ○ That Dr. Lott finds that illegal immigrants are equally likely to commit crimes as US citizens.

- ○ That Dr. Lott finds that illegal immigrants are more likely to commit crimes than US citizens.

>>

Donation: *Excuse* condition

What website visitors will learn

We will post your individual donation decision on our website. Here is an example screenshot of the relevant portion of the website we showed to previous study respondents.

- The page states that **all participants were told about Dr. Lott's study** before deciding whether or not to donate to Fund the Wall
- The page lists individual decisions and whether each participant decided to authorize the donation to Fund the Wall

Study Structure

We surveyed respondents earlier this year. **All participants were told about Dr. Lott's study**. Participants were asked if they wanted to authorize a donation to Fund the Wall, an **organization that seeks to reduce illegal immigration by helping to fund and construct the US-Mexico border wall**.

Donation decisions

- **Alfred Marshall** was told about Dr. Lott's study and **authorized** the donation to Fund the Wall

- **William Jevons** was told about Dr. Lott's study and **did not authorize** the donation to Fund the Wall

- **Carl Menger** was told about Dr. Lott's study and **did not authorize** the

How clear will it be from the website whether or not you chose to donate to Fund the Wall?

- ○ Very clear
- ○ Clear
- ○ Unclear
- ○ Very unclear

People who visit the website **will know** that you were shown the results of Dr. Lott's study before making your donation decision. How clear was this to you from the screenshots we posted above? Please explain in 1-2 sentences.

Reminder: what website visitors will learn

- They will see that Dr. Lott's study finds that illegal immigrants commit more crime than US citizens
- They will learn whether or not you donated to Fund the Wall
- They will learn that **you were informed** about the results of Dr. Lott's study

Donation decision

Would you like to authorize a $1 donation to **Fund the Wall**?

○ Yes, I would like to authorize a $1 donation

○ No, I would not like to authorize a $1 donation

Recall what people will learn when visiting the website:

- They will see that Dr. Lott's study finds that illegal immigrants commit more crime than US citizens
- They will learn whether or not you donated to Fund the Wall
- They will **learn that you were informed** about the results of Dr. Lott's study

Donation: *No Excuse* **condition**

What website visitors will learn

We will post your individual donation decision on our website. Here is an example screenshot of the relevant portion of the website we showed to previous study respondents.

- The page states that **all participants made their decisions before Dr. Lott's study was published**. The page does **not mention** that you were shown an early summary of Dr. Lott's study.
- The page lists individual decisions and whether each participant decided to donate to Fund the Wall.

Study Structure

We surveyed respondents earlier this year **before Dr. Lott's study was published**. Participants were asked if they wanted to authorize a donation to Fund the Wall, an **organization that seeks to reduce illegal immigration by helping to fund and construct the US-Mexico border wall**.

Donation decisions

- **Alfred Marshall** was surveyed before Dr. Lott's study was published and **authorized** the donation to Fund the Wall

- **William Jevons** was surveyed before Dr. Lott's study was published and **did not authorize** the donation to Fund the Wall

- **Carl Menger** was surveyed before Dr. Lott's study was published and **did not authorize** the donation to Fund the Wall

How clear will it be from the website whether or not you chose to donate to Fund the Wall?

○ Very clear

○ Clear

○ Unclear

○ Very unclear

93

People who visit the website **will not know** that you were shown the results of Dr. Lott's study before making your donation decision since the website states that you were surveyed before the study was published and does not mention that you were shown a summary of the study's findings. How clear was this to you from the screenshots we posted above? Please explain in 1-2 sentences.

[››]

Reminder: what website visitors will learn

- They will see that Dr. Lott's study finds that illegal immigrants commit more crime than US citizens
- They will learn whether or not you donated to Fund the Wall
- They will believe **you were not informed** about the results of Dr. Lott's study

Donation decision

Would you like to authorize a $1 donation to **Fund the Wall**?

○ Yes, I would like to authorize a $1 donation

○ No, I would not like to authorize a $1 donation

Recall what people will learn when visiting the website:

- They will see that Dr. Lott's study finds that illegal immigrants commit more crime than US citizens
- They will learn whether or not you donated to Fund the Wall
- They will **not learn that you were informed** about the results of Dr. Lott's study, as the website states that you were surveyed before the study was published and does not mention that you were shown a summary of the study's findings

Donation: control condition

Publicizing the results of our study

As researchers, we believe it is important to communicate our findings about political and social attitudes in Mannheim to the public.

Another group of researchers is working on a related study, and once that study is published in a reputable academic journal, we will post the results from this survey, **including your individual donation decision and the donation decisions of all of the other respondents to this survey who consented on the previous page**, on our website.

We will then work with major news organizations with both a liberal and conservative viewership in Mannheim to publicize our website through newspaper and website articles, and we will also promote our website via Facebook ads to Mannheim residents.

What website visitors will learn

We will post your individual donation decision on our website. Here is an example screenshot of the relevant portion of the website we showed to previous study respondents.

- The page lists individual decisions and whether each participant decided to donate to Fund the Wall.

Study Structure

We surveyed respondents earlier this year. Participants were asked if they wanted to authorize a donation to Fund the Wall, an **organization that seeks to reduce illegal immigration by helping to fund and construct the US-Mexico border wall**.

Donation decisions

- **Alfred Marshall** was surveyed and **authorized** the donation to Fund the Wall

- **William Jevons** was surveyed and **did not authorize** the donation to Fund the Wall

- **Carl Menger** was surveyed and **did not authorize** the donation to Fund the

How clear will it be from the website whether or not you chose to donate to Fund the Wall?

- ○ Very clear
- ○ Clear
- ○ Unclear
- ○ Very unclear

95

Reminder: what website visitors will learn

- They will learn whether or not you donated to Fund the Wall

You

Website visitors

Donation decision

Would you like to authorize a $1 donation to **Fund the Wall**?

◯ Yes, I would like to authorize a $1 donation

◯ No, I would not like to authorize a $1 donation

Recall what people will learn when visiting the website:

- They will learn whether or not you donated to Fund the Wall

Post-outcome measures

Before today, had you taken any previous online surveys that discussed Dr. Lott's study about the crime rates of illegal immigrants?

○ Yes

○ No

If you had to guess, what would you say was the purpose of this study?

F Survey instruments: Experiment 2B

We are a group of **nonpartisan researchers** interested in compiling an accurate and unbiased report about **political and social attitudes in Bergen**.

This survey will take **approximately 5-10 minutes**.

Consent for Participation in a Research Study

Study Title: Political and Social Attitudes

Principal Investigator: Leonardo Bursztyn

IRB Study Number: IRB19-1320

DESCRIPTION: We are nonpartisan researchers doing a research study about political and social attitudes in the United States. The research project will consist of reading information and answering a few short questions. Participation should take approximately 5-10 minutes. We do not include any deceptive questions.

RISKS and BENEFITS: The risks to your participation in this online study are those associated with basic computer tasks, including boredom, fatigue, mild stress, or breach of confidentiality. The only benefit to you, other than survey compensation, is the learning experience from participating in a research study. The benefit to society is the contribution to scientific knowledge.

COMPENSATION: You will be compensated by your survey provider for completing this survey. Partially-completed survey responses will not be compensated.

CONFIDENTIALITY: All data will be stored on password-protected servers and hard drives. We do not ask for any identifying information other than your first and last name, city, and operating system. Unless otherwise explicitly noted, your answers to the questions on our surveys linked to identifying information will never be shared outside of the immediate research team. Any reports and presentations about the findings from this study will not include any other information that could identify you, other than the information listed above. We may share the data we collect in this study with other researchers doing future studies – if we share your data, we will not include any information that could identify you.

SUBJECT'S RIGHTS: Your participation is voluntary. You may stop participating at any time by closing the browser window or the program to withdraw from the study.

For additional questions about this research, you may contact:

Leonardo Bursztyn, The University of Chicago, bursztyn.research@gmail.com.

For questions about your rights as a research participant, you may contact:

The Social & Behavioral Sciences Institutional Review Board, University of Chicago

Phone: (773) 834-7835

E-mail: sbs-irb@uchicago.edu

I have read and understood the above, **I live in Bergen**, and I want to participate in this study.

○ Yes

○ No

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. This means that there are a lot of random answers which compromise the results of research studies. To show that you read our questions carefully, please choose both "Extremely interested" and "Not at all interested" as your answer in the next question. How interested are you in sports?

- ☐ Extremely interested
- ☐ Very interested
- ☐ A little bit interested
- ☐ Almost not interested
- ☐ Not at all interested

››

Are you Spanish, Hispanic, or Latino or none of these?

○ Yes

○ None of these

What is your year of birth?

[▾]

What is your sex?

○ Male

○ Female

In politics, as of today, do you consider yourself a Republican, a Democrat, or an Independent?

○ Republican

○ Democrat

○ Independent

››

Do you strongly support or weakly support the Republican Party?

○ Strongly support

○ Weakly support

››

Do you lean toward the Republican Party or the Democratic Party?

○ Lean toward the Republican Party

○ Lean toward the Democratic Party

››

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)

Which of the following best describes your race or ethnicity?

○ African American/Black

○ Asian/Asian American

○ Caucasian/White

○ Native American, Inuit or Aleut

○ Native Hawaiian/Pacific Islander

○ Other

››

Do Illegal Immigrants Commit More Crime? Preliminary evidence from Dr. Lott's study

Economist Dr. John R. Lott carried out a study on the relationship between illegal immigration and crime. **The study has not been published in an academic journal, but we summarize some of its findings here.**

Dr. Lott's study uses **detailed data spanning over two decades from the Arizona state prison systems**. He finds that **"undocumented immigrants are at least 142% more likely** to be convicted of a crime than other Arizonans. They also tend to commit **more serious crimes**," such as murder and rape.

Moreover, "young undocumented immigrants commit crime at twice the rate of young U.S. citizens." He calculates that "if undocumented immigrants committed crime nationally as they do in Arizona, in 2016 they would have been responsible for over **1,000 more murders, 5,200 rapes, 8,900 robberies, 25,300 aggravated assaults, and 26,900 burglaries**." The vast majority of the illegal immigrants Lott studies are from Mexico.

However, a number of sources (including a researcher affiliated with the Cato Institute, a non-partisan libertarian think tank) have recently challenged some of the study's methods, claiming that errors in analysis invalidate its results. Dr. Lott has responded to this critique with a defense of the study's methods, but the issue remains unresolved.

Do you think Dr. Lott's study will generate widespread media attention?

○ Yes

○ No

On the next screen, you will have the opportunity to authorize a $1 donation to an organization that seeks to reduce illegal immigration into the United States. We will make the donation on your behalf, so **it will not be deducted from your payment**.

››

Donation decision

You now have the opportunity to authorize a $1 donation to **Fund the Wall**.

Fund the Wall is a nonprofit organization that seeks to reduce illegal immigration into the United States by raising money for the construction of the US-Mexico border wall. If you wish, you can choose to authorize a $1 donation to Fund the Wall. The donation **will not be deducted from your payment.**

Before proceeding, please confirm your consent to the following:

I consent to the researchers requesting the following information from my study provider: first and last name and **city. We will not have access to any other personally-identifying information, and we will use the information for study purposes only.**

○ Yes, I consent and would like to proceed with the survey.

○ No, I do not consent and would like to terminate the survey now.

››

F.1 *No Excuse* condition

Publicizing the results of our study

As researchers, we believe it is important to communicate our findings about political and social attitudes in Bergen to the public.

If/when Dr. Lott's study is published in an academic journal, we will post the results from this survey, **including your individual donation decision and the donation decisions of all of the other respondents to this survey who consented on the previous page**, on our website.

We will promote our website via Facebook ads to Bergen residents.

What website visitors will learn

We will post your individual donation decision on our website, which will also include a description of Dr. Lott's study.

- The web page will state that **all participants were surveyed before Dr. Lott's study was published in an academic journal**
- The page lists individual donation decisions: whether or not each participant decided to authorize the donation to Fund the Wall

Donation decision

Would you like to authorize a $1 donation to **Fund the Wall**?

○ Yes, I would like to authorize a $1 donation

○ No, I would not like to authorize a $1 donation

Recall what people will learn when visiting the website:

- The web page will state that **all participants were surveyed before Dr. Lott's study was published in an academic journal**
- The page lists individual donation decisions: whether or not each participant decided to authorize the donation to Fund the Wall

››

F.2 *Excuse* condition

Publicizing the results of our study

As researchers, we believe it is important to communicate our findings about political and social attitudes in Bergen to the public.

If/when Dr. Lott's study is published in an academic journal, we will post the results from this survey, **including your individual donation decision and the donation decisions of all of the other respondents to this survey who consented on the previous page**, on our website.

We will promote our website via Facebook ads to Bergen residents.

What website visitors will learn

We will post your individual donation decision on our website, which will also include a description of Dr. Lott's study.

- The web page will state that **all participants were shown the preliminary findings from Dr. Lott's study** before deciding whether or not to donate to Fund the Wall
- The page lists individual donation decisions: whether or not each participant decided to authorize the donation to Fund the Wall

Donation decision

Would you like to authorize a $1 donation to **Fund the Wall**?

○ Yes, I would like to authorize a $1 donation

○ No, I would not like to authorize a $1 donation

Recall what people will learn when visiting the website:

- The web page will state that **all participants were shown the preliminary findings from Dr. Lott's study** before deciding whether or not to donate to Fund the Wall
- The page lists individual donation decisions: whether or not each participant decided to authorize the donation to Fund the Wall

»

F.3   Post-treatment questions

Before today, had you taken any previous online surveys that discussed Dr. Lott's study about the crime rates of illegal immigrants?

○ Yes

○ No

If you had to guess, what would you say was the purpose of this study?

If you have any feedback on our survey, please leave it below.

››

G Survey instruments: motivating survey

G.1 Instructions: *Excuse* condition

Before proceeding, please re-confirm your consent to the following:

I consent to researchers at Harvard University and University of Chicago accessing the following information from my study provider: **first and last name, and city. There is no need to provide your first and last name, and city in this survey. The survey company we work with has this information already.**

◉ Yes, I consent and would like to proceed with the survey.

○ No, I do not consent and would like to terminate the survey now.

⟫

We are interested in understanding whether you agree with the following statement:

I support a permanent ban on Mexican immigration to protect the US from contagious diseases, such as the coronavirus

As researchers, we believe it is important to communicate whether people in Mannheim support a permanent ban on Mexican immigration.

Therefore, we will post the results from this survey, **including your individual response and the responses of all of the other respondents to this survey who consented on the previous page**, on our study website.

If you indicate that you support a permanent ban on Mexican immigration, our website will read: **"XXX supports a permanent ban on Mexican immigration to protect the US from contagious diseases, such as the coronavirus."** If you indicate that you do not support a permanent ban on Mexican immigration, our website will read: **"XXX does not support a permanent ban on Mexican immigration."**

We will promote our website via Facebook ads targeting both liberals and conservatives residing in Mannheim.

We will post whether you indicate your support for a permanent ban on Mexican immigration on our website. Here is an example screenshot of how the relevant portion of the web page could look like.

We asked people whether they support a permanent ban on Mexican immigration

- Alfred Marshal supports a permanent ban on Mexican immigration to protect the US from contagious diseases, such as the coronavirus
- Wiliam Jevons does not support a permanent ban on Mexican immigration
- Carl Menger supports a permanent ban on Mexican immigration to protect the US from contagious diseases, such as the coronavirus

Do you agree with the following statement?

I support a permanent ban on Mexican immigration to protect the US from contagious diseases, such as the coronavirus

○ Yes

○ No

G.2 Instructions: *No Excuse* condition

Before proceeding, please re-confirm your consent to the following:

I consent to researchers at Harvard University and University of Chicago accessing the following information from my study provider: **first and last name, and city. There is no need to provide your first and last name, and city in this survey. The survey company we work with has this information already.**

○ Yes, I consent and would like to proceed with the survey.

○ No, I do not consent and would like to terminate the survey now.

››

Do you worry that Mexican immigration could make the country more vulnerable to contagious diseases, such as the coronavirus?

○ Yes

○ No

››

We are interested in understanding whether you agree with the following statement:

I support a permanent ban on Mexican immigration

As researchers, we believe it is important to communicate whether people in Mannheim support a permanent ban on Mexican immigration.

Therefore, we will post the results from this survey, **including your individual response and the responses of all of the other respondents to this survey who consented on the previous page**, on our study website.

If you indicate that you support a permanent ban on Mexican immigration, our website will read: **"XXX supports a permanent ban on Mexican immigration."** If you indicate that you do not support a permanent ban on Mexican immigration, our website will read: **"XXX does not support a permanent ban on Mexican immigration."**

We will promote our website via Facebook ads targeting both liberals and conservatives residing in Mannheim.

››

115

We will post whether you indicate your support for a permanent ban on Mexican immigration on our website. Here is an example screenshot of how the relevant portion of the web page could look like.

We asked people whether they support a permanent ban on Mexican immigration

- Alfred Marshal supports a permanent ban on Mexican immigration
- Wiliam Jevons does not support a permanent ban on Mexican immigration
- Carl Menger supports a permanent ban on Mexican immigration

Do you agree with the following statement?

I support a permanent ban on Mexican immigration

○ Yes

○ No

H Survey instruments: Gullibility/Intolerance Dictator Game

H.1 Gullibility condition

We conducted a survey about political and social attitudes in the United States earlier this year. **You have been matched with one of the respondents from that survey.**

Your matched respondent completed a short questionnaire during the survey: **The Gullibility Scale**. This scale measures how easily people are manipulated by evidence from untrustworthy sources. The test is scored from 0 to 100, where 0 means "least gullible" and 100 means "most gullible". Thus, a higher score indicates that your matched respondent is more gullible.

The next screen includes information about how your matched respondent scored on **The Gullibility Scale**.

>>

We conducted a survey about political and social attitudes in the United States earlier this year. **You have been matched with one of the respondents from that survey.**

Your matched respondent completed a short questionnaire during the survey: **The Gullibility Scale**. This scale measures how easily people are manipulated by evidence from untrustworthy sources. The test is scored from 0 to 100, where 0 means "least gullible" and 100 means "most gullible". Thus, a higher score indicates that your matched respondent is more gullible.

The next screen includes information about how your matched respondent scored on **The Gullibility Scale**.

››

H.2 Intolerance condition

We conducted a survey about political and social attitudes in the United States earlier this year. **You have been matched with one of the respondents from that survey.**

Your matched respondent completed a short questionnaire during the survey: **The Foreign Culture Tolerance Scale**. This scale measures whether people are tolerant or intolerant toward foreign values and traditions. The test is scored from 0 to 100, where 0 means "least tolerant" and 100 means "most tolerant". Thus, a higher score indicates that your matched respondent is more tolerant toward foreign values and traditions.

The next screen includes information about how your matched respondent scored on **The Foreign Culture Tolerance Scale**.

››

We conducted a survey about political and social attitudes in the United States earlier this year. **You have been matched with one of the respondents from that survey.**

Your matched respondent completed a short questionnaire during the survey: **The Foreign Culture Tolerance Scale**. This scale measures whether people are tolerant or intolerant toward foreign values and traditions. The test is scored from 0 to 100, where 0 means "least tolerant" and 100 means "most tolerant". Thus, a higher score indicates that your matched respondent is more tolerant toward foreign values and traditions.

The next screen includes information about how your matched respondent scored on **The Foreign Culture Tolerance Scale**.

>>

120