I Have Nothing Against Them, But...*

Leonardo Bursztyn[†] **Ingar Haaland**[‡] **Aakaash Rao**[§] **Christopher Roth**[¶]

March 2020

Abstract

We study the role of commonly known rationales in facilitating the public expression of stigmatized views, such as opposition to a minority group. Consider a setting in which agents may take an anti-minority action for two reasons: because they are intolerant, or because they are gullible and easily persuaded by misinformation. When a rationale to oppose minorities becomes commonly known, tolerant, gullible agents are induced into taking the anti-minority action because they believe the rationale. This gives intolerant, non-gullible agents — who themselves do not believe the rationale — an excuse to take the anti-minority action, allowing them to pool with the tolerant, gullible agents to avoid being seen as intolerant. We examine this mechanism through two related experiments. In the first experiment, subjects guess why participants from a previous survey authorized donations to an anti-immigrant organization. We show that subjects perceive donors who had been exposed to a rationale for holding anti-immigrant views (that immigrants commit more crimes) as less biased against immigrants and more gullible than donors who had not been exposed to the rationale. In the second experiment, subjects choose whether or not to make a publicly-observable donation to an anti-immigrant organization. Subjects who believe that their exposure to the rationale will be publicly observable are substantially more likely to donate than subjects who are only privately exposed to the rationale. We conclude by discussing implications for debunking fake news on social media and understanding populist rhetoric.

Keywords: Social norms; social image; excuses; persuasion; media; propaganda; political attitudes

*We thank Daron Acemoglu, Alberto Alesina, Ben Enke, Ed Glaeser, Daniel Gottlieb, Emir Kamenica, Ross Mattheis, Paulo Natenzon, Nathan Nunn, Doron Ravid, Andrei Shleifer, and seminar participants at Harvard and Warwick for very helpful suggestions. We thank Alison Zhao and particularly Takuma Habu for outstanding research assistance. We gratefully acknowledge financial support from the Pearson Institute for the Study and Resolution of Global Conflicts and the University of Chicago Social Sciences Research Center. The experiments were pre-registered in the AEA RCT Registry with IDs AEARCTR-0005308 and AEARCTR-0005462. The research described in this article was approved by the University of Chicago Social and Behavioral Sciences Institutional Review Board and the Humanities and Social Sciences Research Ethics Committee at the University of Warwick. Replication files are publicly available at `www.github.com/AakaashRao/excuses`.

[†]University of Chicago and NBER, `bursztyn@uchicago.edu`

[‡]University of Bergen and CESifo, `ingar.haaland@uib.no`

[§]Harvard University, `arao@g.harvard.edu`

[¶]University of Warwick, briq, CEPR, CESifo, and CAGE, `christopher.roth@warwick.ac.uk`

1 Introduction

How do anti-minority movements gain traction in societies that stigmatize anti-minority expression? A growing body of evidence suggests that propaganda, political rhetoric, and mass media 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 more anti-Semitic acts in Nazi Germany (Adena et al., 2015). More recently, inflammatory tweets from Donald Trump has led to more anti-Muslim hate crimes (Müller and Schwarz, 2018).

While these studies demonstrate that the media has a causal impact on xenophobic behavior, the underlying mechanisms remain unclear. The most-studied channel is persuasion: for example, one might attribute the growing wave of anti-immigrant rhetoric and violence in the United States to private attitudes toward immigrants becoming more negative.[1] Yet survey evidence suggests quite the opposite. Indeed, both Democrats and Republicans reported feeling, if anything, more *warmly* toward both legal and illegal immigrants in 2018 — after a divisive campaign marked by anti-immigrant rhetoric — than in 2014 (Gonzalez-Barrera and Connor, 2019). Consistent with this observation, recent experimental work finds small or null effects of information on immigration policy preferences (Hopkins et al., 2019; Alesina et al., 2019; Grigorieff et al., 2018), and field experiments consistently show only minimal persuasive effects of political campaigns on behavior (Kalla and Broockman, 2018). Together, the survey and experimental evidence suggests that mechanisms beyond persuasion may be driving trends in public anti-minority expression.

In this paper, we propose an alternative mechanism through which the media might affect public behavior. By creating *common knowledge* about rationales to oppose minorities, whether true or false, the media generates "excuses" for 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 or news network claims that Mexican immigrants bring crime into the country), these people are given an "excuse": they can attribute their position to a belief that Mexicans commit more crime, even if they privately do not believe it is true. The key point is that common knowledge of the excuse opens up a second potential explanation for their anti-Mexican views (i.e. that they are gullible and easily persuaded by misinformation), reducing the extent to which they expect their audience to update about their intolerance.[2]

Populist politicians often use excuses to great effect. US President Donald Trump, for example, built his campaign on a narrative that Mexican immigration causes violent crime. This rationale allowed Trump to pursue aggressive anti-immigration policies while maintaining plausible deniability about these policies' motivations: when asked about Trump's immigration policies, 49 percent of voters answered that they were motivated by a "sincere interest in controlling our borders" (whether

[1] For example, the number of white nationalist hate groups in the United States has grown by 55 percent since January 2017. (Wilson, Jason. "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. (Underwood, Alexia. "Islamophobia in the US: It Goes Way beyond Trump." *Vox News*, April 6, 2018.)

[2] 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.

or not they were the most effective way to accomplish this objective), while 41 percent answered that they are motivated by "racist beliefs" (Snow, 2019). In addition, and often simultaneously, populist politicians can also serve as *suppliers* of excuses, using their platform to generate common knowledge about plausible rationales for otherwise-stigmatized policies and thus emboldening their supporters to publicly voice their positions.[3]

We formalize this intuition with a simple theoretical model. There are two agents: "senders" and "receivers." The agents differ on two dimensions: tolerance (intolerant vs. tolerant toward foreigners) and gullibility (gullible vs. non-gullible). The agents are first exposed to an anti-immigrant rationale that only gullible types believe in. The sender then chooses whether to make a donation to an anti-immigrant organization. The receiver observes the sender's donation decision and makes an inference about whether the sender is tolerant or intolerant. Senders receive expressive utility from making a donation decision consistent with their own tolerance type as well as social utility from making the receiver believe that they are of the same tolerance type.

We model the effect of excuses by varying whether or not the receiver believes that the sender was exposed to the anti-immigrant rationale prior to donating. In the "No Excuse" condition, the receiver incorrectly believes that the sender did *not* learn about the anti-immigrant rationale, and thus believes with certainty that senders who donate are intolerant. Thus, the expected social cost of donating is high and donation rates are low. In contrast, in the "Excuse" condition, the receiver correctly believes that the sender learned about the anti-immigrant rationale, and thus that senders who donate might have done so either because they are intolerant or because they are gullible and were persuaded by the rationale. This reduces the social cost of donating and increases donation rates by allowing intolerant and sophisticated agents to pool with gullible agents — in essence, pretending that their motivation for donating is the anti-immigrant rationale rather than their intolerance.

We examine the role of excuses through two complementary large-scale online experiments. 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.[4] In particular, we inform our respondents that we have matched them with a respondent from another study who took an anti-immigrant action (authorizing a donation to "Fund the Wall"). All respondents in this experiment are informed about a recent study (Lott, 2018), which finds that undocumented immigrants in Arizona commit crimes at substantially higher rates than comparable US citizens.[5] Our key experimental variation is whether our respondents believe that their matched

[3]While we focus in our experiments on crime rates as an excuse, there are a number of common alternative rationales for opposing immigration. One such rationale is that immigrants place an undue burden on American taxpayers — a claim which is demonstrably false, as both legal and illegal immigrants on net contribute *positively* to the tax base. (Campbell, Alexia Fernández. "Undocumented Immigrants Pay Millions of Dollars in State Taxes — Even in the Reddest States." *Vox News*, March 1, 2019.) For example, Ivana Trump (Donald Trump's former wife) explained her position on immigration from Mexico as follows: "I have nothing against Mexicans, but if they [come] here – like this 19-year-old, she's pregnant, she crossed over a wall that's this high...She gives the birth in American hospital, which is for free. The child becomes American automatically. She brings the whole family, she doesn't pay the taxes, she doesn't have a job, she gets the housing, she gets the food stamps. Who's paying? You and me." (Revesz, Rachael. "Donald Trump's Ex-Wife Says She Does Not Want '19- Year-Old Pregnant Mexican Women' Coming to the US." *The Independent*, April 5, 2016.)

[4]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.

[5]The Trump administration has cited this study repeatedly as evidence for the impact of illegal immigration on

respondent knew about the study before making their decision: respondents in the "Excuse" condition are told that their matched respondent knew about the study before making their decision whereas respondents in the "No Excuse" condition are told that their matched respondent was not told about the study.

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 a more structured measure of beliefs: half of the participants make an incentivized guess about their matched respondent's score on a test measuring gullibility, while the other half make an incentivized guess about their matched respondent's score on a test measuring cultural tolerance.

Consistent with our theoretical framework, we find strong treatment effects on both measures. In describing why they believed their matched respondent chose to donate to Fund the Wall, participants matched with a respondent who had no rationale 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 a rationale. 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 from Experiment 1 suggests that publicly known rationales for xenophobic behavior can affect how an audience updates about the underlying motives.

This finding raises the question of whether people strategically use excuses to avoid the social stigma associated with publicly expressing their views. We investigate this question in Experiment 2, in which we recruit a broadly representative sample of 3,728 Republicans and Independents and study whether they are more willing to publicly authorize a donation to Fund the Wall when they have an excuse available.[6]

Our experiment proceeds as follows. We begin by informing participants about the same Lott study, and we give them the opportunity to authorize a $1 donation to Fund the Wall. We tell participants that we will post their individual donation decisions on our website, 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

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 the Cato Institute has challenged the validity of the study (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). In contrast to the study we cite, most work generally find that undocumented immigrants commit crimes at rates similar to or lower than comparable US citizens.

[6]Pilot results suggested that the base rate at which Democrats authorized a donation to Fund the Wall is extremely low, so we focused on Republicans and Independents in this experiment.

in the audience's beliefs ("second-order" persuasion). We hold first- and second-order persuasion fixed across the "Excuse" and "No Excuse" condition by (1) informing participants in both conditions about the study and (2) by making it clear that the website on which their donation decisions will be posted will also contain information about the study, such that all visitors will learn about the study before viewing individual donation decisions. The key treatment varies 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 the 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, and implicitly believe that website visitors will also not be informed — 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 highlights a quantitatively important role of commonly known excuses relative to the effect of persuasion and anticipated persuasion of one's audience. Moreover, the effect is driven by participants who live in majority-Democrat areas, suggesting that participants more strongly require excuses when their audience is likely to disapprove of their actions.

Our paper builds on theoretical literature on the effects of social image concerns on economic and moral decision-making (Bénabou et al., 2018; Bénabou and Tirole, 2006). Most closely related 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 image concerns on political and economic outcomes (including voting, as in DellaVigna et al. 2017; campaign donations, as in Perez-Truglia and Cruces 2017; educational investments, as in Bursztyn et al. 2019a and Bursztyn et al. 2017b; health investments, as in Karing 2018; and political activism, as in Cantoni et al. 2019 and Hager et al. 2019). Kuran (1997) argues that "preference falsification" — expressing a public view distinct from (often opposite of) one's private view in order to conform to perceived social norms — can have dramatic consequences for political equilibria. Bursztyn et al. (2017a) show that updating views about Donald Trump's popularity eliminates the gap between public and private support for an anti-immigrant organization. Similarly, Bursztyn et al. (2020) find that experimentally correcting misperceptions about the acceptability women working outside the home in Saudi Arabia increases the probability that a woman will accept a job outside of the home. 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 thus emphasize the importance of excuses, which can be created by political entrepreneurs and the media, in facilitating xenophobic expression.

Our work also adds to a literature on the effects of media and propaganda on political and eco-

nomic behavior (La Ferrara, 2016; Banerjee et al., 2019; DellaVigna and La Ferrara, 2015; La Ferrara et al., 2012; Bursztyn et al., 2019b). Work on the effects of media on violence (Yanagizawa-Drott, 2014; Müller and Schwarz, 2018), protest participation (Enikolopov et al., forthcoming), and other public outcomes are generally unable to cleanly distinguish between direct persuasion and social image concerns arising from changes in higher-order beliefs. We examine how the media can generate rationales to violate social norms, leading to changes in public behavior even in the absence of persuasive impact.

Finally, our study relates to a literature on moral wiggle room and recent work on "implicit preferences" (Cunningham and de Quidt, 2016). Several lab studies (e.g. Dana et al. 2007; Hamman et al. 2010; Lazear et al. 2012) show that the availability of even weak rationales to behave selfishly (e.g. choosing not to click a button to reveal a matched respondent's payoffs) has substantial effects on behavior. For example, Exley (2016) finds that individuals use risk as a rationale to avoid donating to charitable causes. Because decisions in these settings are anonymous, these findings can be rationalized by a behavioral model of self-signaling, as in Bénabou and Tirole (2011) or, in some cases, by social image concerns vis-a-vis the experimenter. Other work has studied settings where decisions are observable, generating social image concerns. Andreoni and Bernheim (2009) find that increasing the probability that the dictator's choice will be ignored and the recipient allocated an unfavorable amount reduces generosity by giving the dictator "plausible deniability" vis-a-vis the recipient. Conversely, Ariely et al. (2009) show that extrinsic incentives for prosocial behavior can crowd out image motivation (a possibility suggested by Bénabou and Tirole 2006), which can in our framework be interpreted as a "reverse excuse" that decreases the extent to which the audience updates about an agent's prosociality in light of a prosocial action. Our work contributes to this literature by providing evidence on a novel mechanism through which publicly known justifications may affect public behavior. We formalize our proposed mechanism using a two-type equilibrium signaling model, highlighting the use of commonly known rationales to generate excuses in a relevant political economy context.

The remainder of this paper proceeds as follows. In Section 2, we outline a simple model of communication in which the availability of excuses allows xenophobic, sophisticated agents to pool with tolerant, gullible agents when taking a publicly stigmatized action. 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 a second online experiment demonstrating that commonly known rationales increase xenophobic expression. Section 5 discusses robustness of our experimental findings to attrition and experimenter demand. Section 6 discusses implications of our results for understanding populist rhetoric and organized anti-minority violence. We discuss policy implications and conclude in Section 7.

2 Theoretical Framework

To organize thoughts and motivate our experimental designs, we present a simple model of communication that formalizes the strategic implications of a publicly known rationale for xenophobic behavior.

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 gullible and easily persuaded by misinformation whereas others are non-gullible. 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 non-gullible 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 news story. The sender can choose either to donate to an anti-immigrant organization ($a = 1$) or not to donate ($a = 0$). The receiver observes the sender's donation decision and makes an inference about whether the sender is tolerant or intolerant.

The gullible sender is non-strategic, with actions characterized as follows: in the absence of viewing anti-immigrant information, the tolerant-gullible sender chooses not to donate, while the intolerant-gullible sender donates. However, once exposed to anti-immigrant information, the tolerant-gullible sender is persuaded and induced to donate; the intolerant-gullible sender continues to donate.

The non-gullible 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 $\overline{b}$, while when the receiver believes with certainty that the sender is of the opposite tolerance type, the sender receives social utility $\underline{b}$, with $\overline{b} > \underline{b}$. 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 $b_0 := p\overline{b} + (1-p)\underline{b}$, while the sender's social utility from being perceived as intolerant with certainty is given by $b_1 := p\underline{b} + (1-p)\overline{b}$.

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.[7]

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

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

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

Proposition 1. *Non-gullible senders' optimal actions are as follows*[8]*:*

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

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

[7] 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 $b_0 := \overline{b}$ and $b_1 := \underline{b}$ and the remainder of this section would look identical under this alternative assumption.

[8] We assume that the sender does not donate when she is indifferent between donating and not donating; however, the results in the section do not depend on this assumption.

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 the anti-immigrant news story 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 donates 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 news story prior to choosing her action. Thus, the receiver no longer knows with certainty that a sender who donates is intolerant, since he knows he may be matched with a tolerant-gullible sender who was persuaded by the news story to donate. 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).[9]

The fact that the tolerant-nongullible sender does not donate in either game is immediate, since both social and expressive utility are strictly greater when the tolerant-nongullible sender does not donate than when she donates.[10] When expressive utility v is small relative to social utility, the intolerant-nongullible sender does not donate either in the Excuse game or the No Excuse game because the social image costs of donating outweigh the expressive benefits. In contrast, when expressive utility v is large relative to social utility, the intolerant-nongullible sender donates 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-nongullible sender does not donate under the No Excuse game but donates under the Excuse game, assuming that the share of gullible agents is sufficiently large to allow intolerant-nongullible agents to pool with tolerant-gullible agents. We formalize this claim in Proposition 2, which we prove in Appendix A.

Proposition 2. *Suppose that*

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

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-nongullible sender does not donate in either game, while the intolerant-nongullible sender donates 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.

[9] In our model, gullible and non-gullible receivers are identical. In particular, tolerant-gullible receivers who are persuaded by the anti-immigrant organization still judge intolerant senders in the same manner as tolerant-nongullible receivers, capturing the intuition that people care about the *motivations* behind others' actions. Moreover, gullible receivers still use Bayes' rule to make inferences about the sender's motivations. We could alternatively model gullible and non-gullible receivers differently, such that gullible 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.

[10] 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 $p > 0.5$.

Corollary 1. *In the equilibria as in Proposition 2, the receiver's posterior belief that a sender who donates 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 donates is non-gullible 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-gullible.

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

3 Experiment 1: Excuses and Type Inference

We first examine how the availability of a rationale changes how an audience interprets the decision to donate to Fund the Wall, thus evaluating our predictions about type inference (Corollary 1).[11] 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.[12] As we show in Experiment 2, public behavior among people who live in counties with a lower Republican vote share is substantially more elastic to the availability of an excuse than public behavior among those who live in more Republican counties, suggesting that non-Republicans are indeed the relevant audience to consider when studying how excuses affect inference.

3.1 Sample

We conducted the experiment with Luc.id, a widely used online panel provider (Wood and Porter, 2019). We recruited a sample of 3,047 Democrats in February 2020.[13] 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 A1) and well-balanced on observables across treatment arms (Appendix Table A2). All experimental procedures and analyses were pre-registered in the AEA RCT Registry.

[11] All survey instruments are available in Appendix C.

[12] 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).

[13] 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 nearly identical to the instrument used in the main experiment. We report both unpooled and pooled specifications.

3.2 Experimental design

We tell all respondents about a recent study (Lott, 2018) which finds that 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 45% more likely to be gang members than U.S. citizens."[14] 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.[15]

We then truthfully 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 nonprofit organization that seeks to reduce illegal immigration into the United States by helping to fund and construct the US-Mexico border wall. We make it clear that the respondents from this survey knew that their donation decision would be posted on our study website. 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 learn 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 learned 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". Participants in the "tolerance" condition are truthfully 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 truthfully 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.[16] 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.[17]

[14]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).

[15]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).

[16]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 between design in order to minimize experimenter demand effects and to avoid order effects (Charness et al., 2013).

[17]The previous study respondents with whom Experiment 1 subjects were matched completed a survey very similar in structure to our Experiment 2 survey, but the two surveys were not precisely the same. In particular, it was

Figure 1 outlines the structure of Experiment 1.

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 priming them on the particular dimensions in which we are interested. 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.[18]

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.[19] 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

important that Experiment 1 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 2 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 1, and we matched Experiment 1 subjects with participants from this auxiliary survey.

[18] 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 more depth in Section 5.1.

[19] 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.

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.[20] These are substantial effect sizes, which highlight that the availability of a rationale strongly changes people's inference about their matched respondent's motives. Table 1 displays results in regression form and demonstrates 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, we show in Table A3 that they are highly correlated: on average, a respondent who uses a word related to intolerance (gullibility) when describing the matched respondent's motive rates the matched respondent is half a standard deviation more intolerant (gullible) than a respondent who does not use such a word.

Taken together, our evidence suggests that when judging others' motives, people believe that those who donated with an excuse are more gullible and less intolerant than those who donated without an excuse. Our results thus confirm the first key prediction of our signaling model.

4 Experiment 2: Excuses and Xenophobic Expression

Our results in Experiment 1 beg the question of whether agents strategically use excuses to disguise their intolerance toward immigrants. In this section, we present the results of a second pre-registered experiment testing this hypothesis. Because pilot results suggested that the base rate at which Democrats authorized a donation to Fund the Wall was extremely low, we focused in this experiment on Republicans and Independents.

4.1 Sample

We once again worked with Luc.id to recruit 3728 self-reported Republicans and Independents in a pre-specified experiment in January 2020. 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. Participants were directed to our survey on the online platform *Qualtrics*. As before, 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.[21] Our sample of respondents is broadly representative of Independents and Republicans in the United States (Appendix Table A5) and well-balanced in terms of observables across treatment arms (Appendix Table A6). As in Experiment 1, we pre-registered all experimental procedures and analyses.

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

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

4.2 Experimental design

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. Respondents in the Excuse and No Excuse condition receive the same information as respondents from Experiment 1: they are told about the Lott study, which finds that immigrants commit more crimes and more serious crimes than US citizens, and about the fact that the study has been challenged by a variety of sources. 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, and the respondents will have the opportunity to authorize a $1 donation to this organization. In practice, we randomized almost all of our respondents to Fund the Wall.

Visibility manipulation Our goal is non-deceptively ensure participants believe their individual donation decision will be 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. We inform respondents that we will post the results from the survey, including their "individual donation decision," on our study website. However, despite the fact that all participants who completed the survey consented to us accessing their full names, we decided not to post names in order to avoid potentially compromising participants' privacy. We instead posted anonymized study IDs alongside donation decisions, thus avoiding deception given that "individual donation decisions" were still posted.

We also inform our respondents that "we believe it is important to communicate our findings about political and social attitudes in [City of respondent] to the public".[22] 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.[23]

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. 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

[22] We used participants' IP address to capture and display their current location (i.e. their city).

[23] In practice, we do so by sending our working paper to news organizations and by publicizing the website via Facebook Advertisements.

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 Dr Lott's study.

The key experimental treatment thus cleanly varies the availability of an excuse for donating. In the Excuse treatment, 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 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, this would be deceptive. Instead, we exploit the fact that Lott's study has not yet been published (i.e. it is currently a working paper on SSRN). 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 "Website visitors will believe that you did not know about Dr. Lott's study, 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". Respondents in this condition thus believe that their audience will believe that they (respondents) had no information excusing their decision to donate to fund the border wall. As forming correct higher-order beliefs may be cognitively challenging, we include an illustration of their audience's information set tailored in each treatment condition (Figure A1).

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.

Figure 3 outlines the structure of Experiment 2.

4.3 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:

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

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 1. 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 compared to 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. These effects are robust to the inclusion of control variables and are almost identical in our pre-specified main study and a pilot study. However, 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 small or null effects on behavior and stated preferences (Hopkins et al., 2019; Alesina et al., 2019; Grigorieff et al., 2018). 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." However, 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 respondent's own private beliefs, 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. 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. In particular, the excuse effect is significantly larger for people from counties with a lower Republican vote share, consistent with our model's prediction that the excuse effect should be larger when the share of agents who privately approve of the action is smaller.

5 Robustness

5.1 Demand effects

One concern with our estimated treatment effects could be that respondents across the different treatment conditions hold different beliefs about the experimenter's expectations, and that those beliefs drive the estimated treatment effects. Despite recent research highlighting 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 conduct a number of additional exercises to address the potential for demand effects to bias our findings.

Perceived purpose: machine learning We measured respondent's beliefs about the purpose of the experiment at the end of both Experiment 1 and Experiment 2 using an open-ended question: "If you had to guess, what was the purpose of this study?". To examine whether people in the different treatment conditions hold different beliefs about the purpose of the study, we employ a novel approach applying machine learning techniques to these text responses. In particular, we train a Support Vector 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 in Experiment 1 better than chance (Table A4), suggesting that the treatment does not significantly respondent's perceptions about the purpose of the study. Similarly, we cannot predict treatment status better than chance when distinguishing between the Excuse and No Excuse conditions in Experiment 2 (Table A9). However, we can predict assignment to the Control condition substantially better than chance (Table A9), 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 Dr Lott's 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 Excuse and No Excuse if such differences were present.

Perceived purpose: hand-coding We also hired two independent research assistants to hand-code the responses to the open-ended purpose question in Experiment 2. Table A8 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 A8 also shows that on almost all of the dimensions we code, beliefs about the purpose of the study do not significantly differ between Excuse and No Excuse. 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 different 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 relative to 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 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).

Experiment 2 heterogeneity Finally, heterogeneous treatment effects by the county-level Republican vote share are inconsistent with experimenter demand effects driving our findings. In particular, for demand effects to bias our estimates upward, we would require both that (1) respondents in counties with a lower Republican vote share are substantially more affected by experimenter demand effects than respondents in counties with a higher Republican vote share, and (2) that respondents fail to report differences in perceived purpose at the end of the study, a contingency we view as implausible.

5.2 Differential attrition

Could patterns of differential attrition explain the estimated treatment effects in our data? 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 B in the Appendix). Similarly, in Experiment 2, we find no differential attrition among respondents in the Excuse versus No Excuse condition ($p = 0.47$), and once again, there is no evidence of differential attrition between subgroups (Table B 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 we attribute to the greater survey length of the Excuse and No Excuse versions of the survey. This does not our estimates of the main effect of interest (No Excuse vs. Excuse), but may slightly bias the benchmark (Control vs. No Excuse).

6 Discussion

In this section, we apply our theoretical framework to understand historical and present-day anti-minority expression.

6.1 Populist Rhetoric and Dog-Whistling

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. As Norris and Inglehart (2019) argue:

> "Populist rhetoric tells a simple story about the silent majority of ordinary, hard-working people... Populism rejects the legitimacy of authority derived from scientific evidence, book learning, and reasoned deliberation. Instead, the discourse celebrates the authenticity of direct experience ('Believe me'), mass opinions ('Many people say...'), and quick applause lines ('Build the Wall')."

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.[24] Through our approach, we interpret these appeals to anonymous authority in two ways. First, populists generates an excuse for their own rhetoric: if their comments backfire, they can distance themselves by claiming they only meant to report on others' beliefs, not express their private opinions.[25] For example, upon facing criticism for citing a conspiracy theory concerning the suicide of Vince Foster, a friend of Hillary and Bill Clinton, Trump stated:

> "Somebody asked me the question the other day, and I said that a lot of people are very skeptical as to what happened and how he died. I know nothing about it...I don't think that it's something that should be part of the campaign."[26]

More subtly, such rhetoric also generates the perception of *common knowledge* of the excuse: by implying that fringe conspiracy theories are known to a large group of people (and appearing to endorse the theory themselves), populists' comments positively update their audience's perceptions as to the likelihood that others will believe that they themselves were convinced by the story, and through this channel, the perceived effectiveness of the excuse.[27]

A closely related phenomenon 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). The term was first introduced in its modern form to describe the rhetoric of Australian Prime Minister John Howard over the course of his tenure (1996-2007) and re-election campaigns, during which opponents accused him of employing terms such as "un-Australian" and "mainstream" to signal racist positions while maintaining plausible deniability (Soutphommasane, 2009); the term has also been used retrospectively to describe the Republican Party's "Southern strategy" to win white support in the South by appealing to racist tensions during the 1970s and 80s (Haney-López, 2014). In a 1981 interview, Republican strategist and Republican National Committee chairman Lee Atwater described the strategy as follows:

> "Y'all don't quote me on this. 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

[24] Johnson, Jenna. "'A Lot of People Are Saying...': How Trump Spreads Conspiracies and Innuendos." *Washington Post*, June 13, 2016. As of December 1, 2019, Donald Trump had used the phrases *"people say"* or *"people are saying"* a combined 591 times in speeches, interviews, and Tweets since launching his presidential campaign in June 2015 (Factbase, n.d.). Johnson (see footnote 25) highlights a number of examples, including comments about Barack Obama's religion, Ted Cruz' birthplace, the Iran nuclear deal, extremist Islamic terrorism, and the existence of a "deep state" dedicated to undermining the Trump administration.

[25] This is an example of a *motte-and-bailey* argument (Shackel, 2005), through which a speaker can make a controversial, difficult-to-defend statement (the "bailey") and then retreat to a less controversial and easier-to-defend position when challenged (the "motte"), claiming that the audience had misinterpreted his or her original statement. Such rhetorical devices can thus be used to generate excuses for making false, extreme, or misleading remarks.

[26] Johnson, Jenna. "'A Lot of People Are Saying...': How Trump Spreads Conspiracies and Innuendoes." *Washington Post*, June 13, 2016.

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

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)

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.[28]

6.2 The Eichmann Trial

Adolf Eichmann, one of the primary organizers of the Final Solution, was captured by Mossad agents in 1960 and brought to Jerusalem to stand trial for war crimes, crimes against humanity, and crimes against the Jewish people. As political theorist Hannah Arendt noted in her famous account of the trial *Eichmann in Jerusalem*: Eichmann's testimony betrayed "no case of insane hatred of Jews, of fanatical anti-Semitism or indoctrination of any kind. He *personally* never had anything whatever against Jews" (Arendt, 2006). Instead, Eichmann maintained throughout the trial that he was only following orders and thus bore no culpability for the consequences thereof.[29] As he wrote in his appeal:

> "I detest as the greatest of crimes the horrors which were perpetrated against the Jews and think it right that the initiators of these terrible deeds will stand trial before the law now and in the future. Notwithstanding, there is a need to draw a line between the leaders responsible and the people like me forced to serve as mere instruments in the hands of the leaders. I was not a responsible leader, and as such do not feel myself guilty." [30]

Yet more recently, a number of scholars (Lozowick, 2003; Cesarani, 2006; Lipstadt, 2011) have argued that Eichmann's trial affect — that he had nothing against Jews *per se* and that he was a mere low-level bureaucrat — was deeply at odds with his true beliefs. As Stangneth (2015) notes in her seminal analysis of the Sassen tapes, a series of interviews between Eichmann and former Nazi journalist Willem Sassen in 1957: "He was anything but the unthinking "functionary," the robotic desk-murderer that Arendt had made him out to be. As the Sassen tapes and other testimony reveal, Eichmann was a true believer in the Nazi cause, an impassioned anti-Semite who acted out of deep-seated ideological conviction." (Wolin, 2016). Indeed, as Eichmann said during a speech to fellow Nazi fugitives in Buenos Aires in 1955, just five years before his trial:

> "I regret nothing... My inner self bridles at the thought that we did something wrong. No, I say to you quite honestly, had we killed 10.3 million Jews out of the 10.3 million we had in our sights, I would be quite satisfied, and would say that we annihilated an enemy." (Wolin, 2016)

[28] 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 (2011).

[29] This defense was so common among accused former Nazis that it became known as the "Nuremberg defense". Rudolf Hoess, commandant at Auschwitz-Birkenau, defended his actions similarly: ""Don't you see, we SS men were not supposed to think about these things: it never even occurred to us." (Toland, 2014).

[30] The New York Times. "Letter by Adolf Eichmann to President Yitzhak Ben-Zvi of Israel." *The New York Times*, January 27, 2016.

In his forward to Arendt's *Eichmann in Jerusalem*, Amos Elon writes that "evil, as [Arendt] saw it, need not be committed only by demonic monsters but — with disastrous effect — by morons and imbeciles as well." Holding fixed his action, Eichmann's defense rested on *pooling with these "morons and imbeciles"* — or the bureaucratic, law-abiding functionaries — in order to disguise his true racist (and stigmatized) beliefs. While ultimately unsuccessful in persuading his accusers to commute his sentence, Eichmann's testimony succeeded for many decades in shielding his legacy, portraying him as a gullible and easily-manipulated bureaucrat rather than a fervently anti-Semitic mass murderer — a narrative that has only recently begun to unravel.

7 Conclusion

Motivated by the recent wave of anti-immigrant rhetoric and policy in the United States and Western Europe, we study the effect of a popular rationale — that immigrants commit crimes at vastly higher rates than citizens — on respondent's willingness to donate to an anti-immigrant organization. We present a simple theoretical framework in which the availability of a rationale affects the beliefs a receiver forms about the motives underlying xenophobic expression. In two related experiments, we show that the availability of this rationale decreases the extent to which those who disapprove of the organization believe that the donor is intolerant, and that rationales can thus serve as excuses that increase the prevalence of stigmatized behavior in equilibrium. Finally, we use our approach to examine notable incidents of anti-minority expression, both historically and in the present.[31]

Our findings can also inform the debate about the influence of fake news on society and politics. While studies suggest that the persuasive effect of fake news is not very strong (Nyhan, 2018), our study points to an alternative mechanism through which fake news can affect public expression. That is, our evidence points to the importance of a "persuasion multiplier": fake news that plausibly persuades a small subset of the population 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

[31] Our logic has policy-relevant implications in other contexts. For example, negative stigma surrounding educational investment in some low-income and minority communities (Austen-Smith and Fryer, 2005; Bursztyn et al., 2019a) might be overcome by offering financial incentives (Levitt et al., 2016) to students, providing them with an excuse vis-a-vis their peers for exerting effort in school. Another policy-relevant example is climate change denial: to the extent that the denial of unambiguous facts is stigmatized, people can use the extremely small minority of scientific studies that find mixed or no evidence for anthropogenic climate change as an excuse to oppose environmental protection policies.

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.[32] 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 pre-post 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 scale up the feature to their entire user-base.

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 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?

[32] 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, Brenden. "Why the Fact-Checking at Facebook Needs to Be Checked." *The New York Times*, October 23, 2017.)

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*, November 2015, *130* (4), 1885–1939.

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

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

Arendt, Hannah, *Eichmann in Jerusalem: A Report on the Banality of Evil* Penguin Classics, New York, N.Y: Penguin Books, 2006. OCLC: ocm65198074.

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

Austen-Smith, David and Roland G. Fryer, "An Economic Analysis of "Acting White"," *The Quarterly Journal of Economics*, May 2005, *120* (2), 551–583.

Banerjee, Abhijit, Eliana La Ferrara, and Victor H Orozco-Olvera, "The Entertaining Way to Behavioral Change: Fighting HIV with MTV," Working Paper 26096, National Bureau of Economic Research July 2019.

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

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

_ **, Armin Falk, and Jean Tirole**, "Narratives, imperatives, and moral reasoning," Working Paper 24798, National Bureau of Economic Research July 2018.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," Working Paper, The University of Chicago March 2020.

_ **, Georgy Egorov, and Robert Jensen**, "Cool to Be Smart or Smart to Be Cool? Understanding Peer Pressure in Education," *The Review of Economic Studies*, July 2019, *86* (4), 1487–1526.

_ **,** _ **, and Stefano Fiorin**, "From extreme to mainstream: How social norms unravel," Working Paper 23415, National Bureau of Economic Research May 2017.

_ **,** _ **, Ruben Enikolopov, and Maria Petrova**, "Social media and xenophobia: evidence from Russia," Working Paper 26567, National Bureau of Economic Research December 2019.

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

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement," *The Quarterly Journal of Economics*, May 2019, *134* (2), 1021–1077.

Cesarani, David, *Becoming Eichmann: Rethinking the Life, Crimes, and Trial of a "Desk Murderer"*, Da Capo Press, April 2006.

Charness, Gary, Uri Gneezy, and Alex Imas, "Experimental Methods: Eliciting Risk Preferences," *Journal of Economic Behavior & Organization*, March 2013, *87*, 43–51.

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

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*, October 2007, *33* (1), 67–80.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, November 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, January 2015, pp. 723–768.

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

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova, "Social Media and Protest Participation: Evidence from Russia," *Econometrica*, forthcoming.

_ **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, January 2015, pp. 687–700.

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

Factbase, "The complete collection of Donald Trump interviews, speeches, tweets – including deleted tweets," https://factba.se. Online; accessed 16 March 2020.

Ferrara, Eliana La, "Mass media and social change: Can we use television to fight poverty?," *Journal of the European Economic Association*, July 2016, *14* (4), 791–827.

_ , **Alberto Chong, and Suzanne Duryea**, "Soap operas and fertility: Evidence from Brazil," *American Economic Journal: Applied Economics*, October 2012, *4* (4), 1–31.

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

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*, March 2019.

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

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal, "Does Information Change Attitudes Towards Immigrants?," Working Paper 2768187, Social Science Research Network July 2018.

Hager, Anselm, Lukas Hensel, Johannes Hermle, and Christopher Roth, "Political Activists as Free-Riders: Evidence from a Natural Field Experiment," Working Paper 12759, IZA Institute for Labor Economics November 2019.

Hamman, John R., George Loewenstein, and Roberto A. Weber, "Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship," *American Economic Review*, September 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*, January 2019, *81* (1), 315–320.

Kalla, Joshua L and David E Broockman, "The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments," *American Political Science Review*, 2018, *112* (1), 148–166.

Karing, Anne, "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone," Working Paper November 2018. Retrieved from https://pdfs.semanticscholar.org/5fc4/432c68b7baf40d37e281a0e2fe122d035138.pdf.

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

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

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

Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff, "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance," *American Economic Journal: Economic Policy*, November 2016, *8* (4), 183–219.

Lipstadt, Deborah E., *The Eichmann Trial*, New York: Schocken, March 2011.

Lott, John R., "Undocumented Immigrants, U.S. Citizens, and Convicted Criminals in Arizona," Working Paper 3099992, Social Science Research Network February 2018.

Lozowick, Yaacov, *Hitler's Bureaucrats: The Nazi Security Police and the Banality of Evil*, Continuum, June 2003.

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

Müller, Karsten and Carlo Schwarz, "Making America Hate Again? Twitter and Hate Crime Under Trump," Working Paper 3149103, Social Science Research Network March 2018.

Norris, Pippa and Ronald Inglehart, *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*, New York, NY: Cambridge University Press, February 2019.

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

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

Ousey, Graham C. and Charis E. Kubrin, "Immigration and Crime: Assessing a Contentious Issue," *Annual Review of Criminology*, January 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*, August 2017, *125* (4), 1208–1243.

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

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, April 2019.

Shackel, Nicholas, "The Vacuity of Postmodernist Methodology," *Metaphilosophy*, April 2005, *36* (3), 295–320.

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

Soutphommasane, Tim, *Reclaiming Patriotism: Nation-Building for Australian Progressives* Australian Encounters, New York: Cambridge University Press, October 2009.

Stangneth, Bettina, *Eichmann Before Jerusalem: The Unexamined Life of a Mass Murderer*, translation edition ed., New York: Vintage, August 2015.

Toland, John, *Adolf Hitler: The Definitive Biography*, Knopf Doubleday Publishing Group, September 2014.

Wolin, Richard, "Richard H. King. Arendt and America.," *The American Historical Review*, October 2016, *121* (4), 1244–1246.

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

Yanagizawa-Drott, David, "Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, November 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

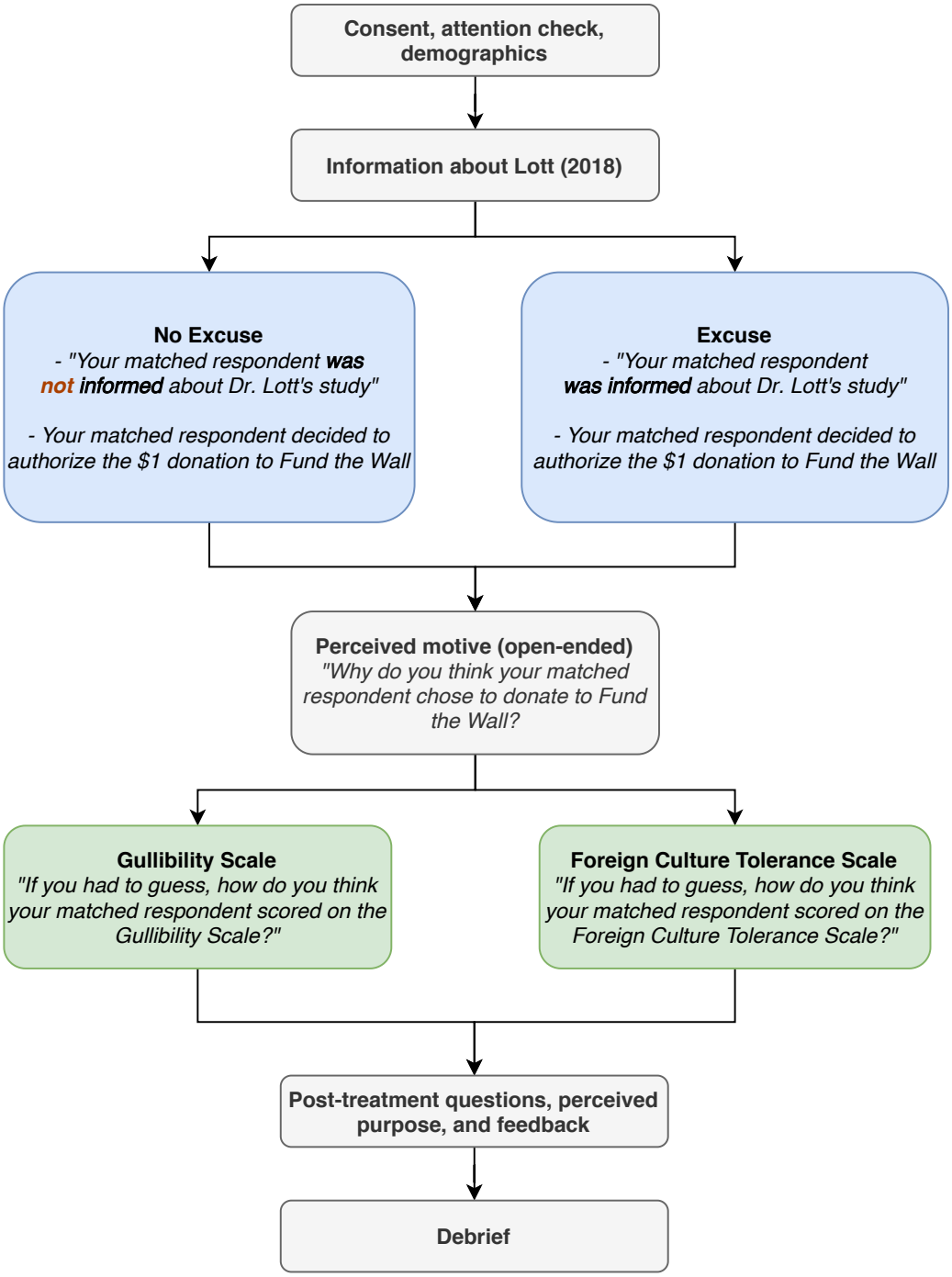


Figure 1: Experiment 1 – structure of design

Tables

Table 1: Experiment 1: Inferred donation motives

	Dependent variable:					
	Inference about partner's donation motive					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *Probability of using word relating to bias*						
Excuse	-0.070^{***}	-0.068^{***}	-0.068^{***}	-0.072^{***}	-0.072^{***}	-0.072^{***}
	(0.012)	(0.012)	(0.012)	(0.009)	(0.009)	(0.009)
Constant	0.172^{***}	0.310^{***}	0.291^{***}	0.169^{***}	0.288^{***}	0.268^{***}
	(0.009)	(0.071)	(0.072)	(0.007)	(0.055)	(0.055)
Observations	3,047	3,047	3,047	5,065	5,065	5,065
R^2	0.010	0.028	0.029	0.011	0.024	0.025
Adjusted R^2	0.010	0.024	0.024	0.011	0.021	0.022
Panel B: *Probability of using word relating to gullibility*						
Excuse	0.031^{***}	0.032^{***}	0.032^{***}	0.029^{***}	0.029^{***}	0.029^{***}
	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)
Constant	0.069^{***}	0.112^{*}	0.118^{**}	0.068^{***}	0.111^{**}	0.113^{**}
	(0.007)	(0.058)	(0.058)	(0.005)	(0.045)	(0.045)
Observations	3,047	3,047	3,047	5,065	5,065	5,065
R^2	0.003	0.010	0.011	0.003	0.009	0.009
Adjusted R^2	0.003	0.006	0.006	0.003	0.006	0.006
Demographic controls	No	Yes	Yes	No	Yes	Yes
Partisan affiliation controls	No	No	Yes	No	No	Yes
Include pilot data	No	No	No	Yes	Yes	Yes

Notes: The dependent variable in Panel A is the probability that the respondent uses a word relating to bias when describing why he or she thinks the matched respondent donated to Fund the Wall, while the dependent variable in Panel B is the probability that the respondent uses a word relating to gullibility in response to the same question. Columns 1-3 report results estimated on the sample from the main experiment, while Columns 4-6 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. Partisan affiliation controls include a dummy for strong Democrats, with weak Democrats as the reference category. Robust standard errors are reported.

Table 2: Experiment 1: Inferred bias and gullibility scores

	Dependent variable:					
	Inference about partner's score					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *Bias (z-score)*						
Excuse	-0.134^{***}	-0.133^{***}	-0.133^{***}	-0.149^{***}	-0.152^{***}	-0.152^{***}
	(0.051)	(0.051)	(0.051)	(0.040)	(0.039)	(0.039)
Constant	0.058	−0.159	−0.174	0.074^{***}	0.047	0.006
	(0.036)	(0.282)	(0.285)	(0.028)	(0.224)	(0.226)
Observations	1,524	1,524	1,524	2,532	2,532	2,532
R^2	0.004	0.038	0.038	0.006	0.037	0.037
Adjusted R^2	0.004	0.030	0.029	0.005	0.032	0.032
Panel B: *Gullibility (z-score)*						
Excuse	0.321^{***}	0.312^{***}	0.310^{***}	0.313^{***}	0.316^{***}	0.317^{***}
	(0.050)	(0.050)	(0.050)	(0.039)	(0.039)	(0.039)
Constant	-0.155^{***}	0.016	−0.100	-0.159^{***}	−0.116	−0.224
	(0.036)	(0.299)	(0.301)	(0.028)	(0.231)	(0.233)
Observations	1,523	1,523	1,523	2,533	2,533	2,533
R^2	0.026	0.060	0.065	0.025	0.055	0.059
Adjusted R^2	0.025	0.052	0.056	0.024	0.050	0.053
Demographic controls	No	Yes	Yes	No	Yes	Yes
Partisan affiliation controls	No	No	Yes	No	No	Yes
Include pilot data	No	No	No	Yes	Yes	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. Columns 1-3 report results estimated on the sample from the main experiment, while Columns 4-6 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. The partisan affiliation control is an indicator that takes value 1 if the respondent self-reports being a strong Democrat, with weak Democrat as the reference category. 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	0.064***	0.060***	0.065***	0.068***	0.067***	0.074***
	(0.020)	(0.020)	(0.018)	(0.018)	(0.017)	(0.016)
Control	−0.0003	−0.007	−0.004	−0.008	−0.011	−0.00004
	(0.020)	(0.019)	(0.018)	(0.019)	(0.018)	(0.017)
p-value (Excuse = Control)	0.0014	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Demographic controls	No	Yes	Yes	No	Yes	Yes
Partisan affiliation controls	No	No	Yes	No	No	Yes
Include pilot data	No	No	No	Yes	Yes	Yes
DV mean	0.489	0.489	0.489	0.498	0.498	0.498
DV std. dev.	0.500	0.500	0.500	0.500	0.500	0.500
Observations	3,728	3,728	3,728	4,444	4,432	4,432
R^2	0.004	0.060	0.188	0.005	0.061	0.198
Adjusted R^2	0.003	0.056	0.184	0.004	0.058	0.195

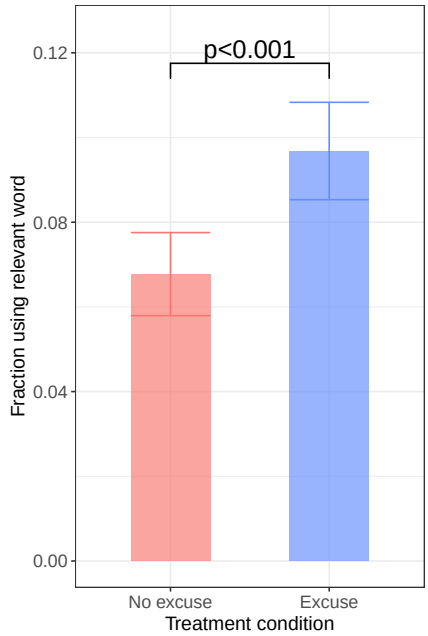
Notes: Columns 1-3 report results estimated on the sample from the main experiment, while Columns 4-6 pool the sample from the main experiment with the sample from the pilot. Demographic controls include age, age squared, a set of race indicators, a Hispanic indicator, a male indicator, and a set of education indicators. Partisan affiliation controls include dummies for strong Republican, weak Republican, Republican-leaning Independent, and Democrat-leaning Independent. Robust standard errors are reported.

Table 4: Experiment 2: County heterogeneity

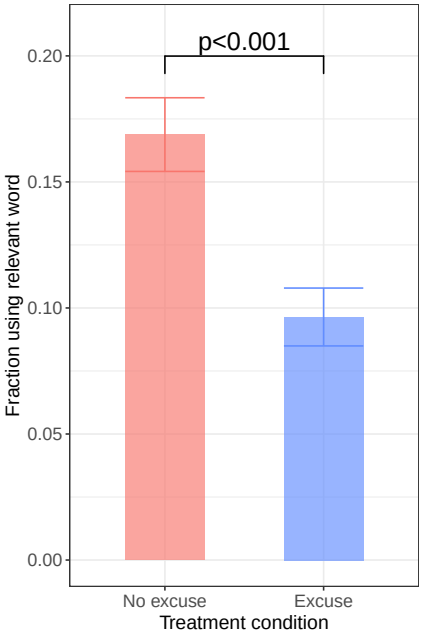
	Dependent variable:					
	Donated to Fund the Wall					
	All		Republicans		Independents	
	(1)	(2)	(3)	(4)	(5)	(6)
Excuse	0.062***	0.070***	0.061**	0.069***	0.061**	0.071***
	(0.019)	(0.016)	(0.029)	(0.025)	(0.024)	(0.022)
Excuse × County Republican vote share	−0.037**	−0.037**	−0.011	−0.013	−0.059**	−0.059***
	(0.018)	(0.016)	(0.029)	(0.025)	(0.024)	(0.022)
Control	−0.005	−0.002	−0.006	−0.004	0.005	0.008
	(0.018)	(0.017)	(0.028)	(0.026)	(0.024)	(0.023)
Control × County Republican vote share	0.019	0.017	0.018	0.012	0.021	0.022
	(0.013)	(0.012)	(0.020)	(0.017)	(0.018)	(0.016)
County Republican vote share	0.010	0.011	−0.008	−0.006	0.026	0.023
	(0.018)	(0.017)	(0.029)	(0.027)	(0.024)	(0.023)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Partisan affiliation controls	Yes	Yes	Yes	Yes	Yes	Yes
Include pilot data	No	Yes	No	Yes	No	Yes
DV mean	0.489	0.498	0.489	0.498	0.489	0.498
DV std. dev.	0.500	0.500	0.500	0.500	0.500	0.500
Observations	3,631	4,315	1,551	1,920	2,080	2,395
R^2	0.192	0.203	0.071	0.073	0.142	0.156
Adjusted R^2	0.188	0.199	0.060	0.064	0.134	0.150

Notes: Republican vote share 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. Partisan affiliation controls include dummies for strong Republican, weak Republican, Republican-leaning Independent, and Democrat-leaning Independent. Robust standard errors are reported.

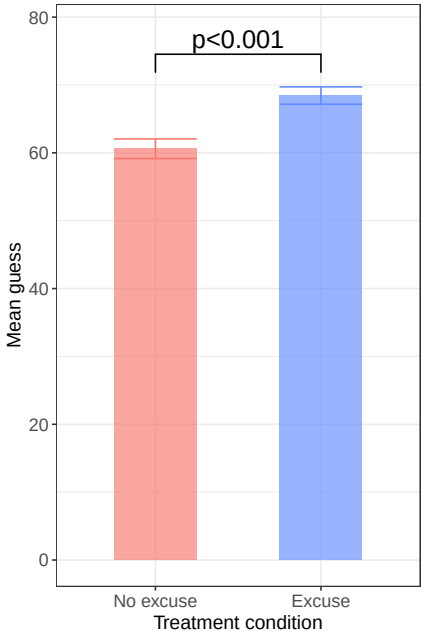
Figure 2: Experiment 1: Type inference based on text analysis and scales



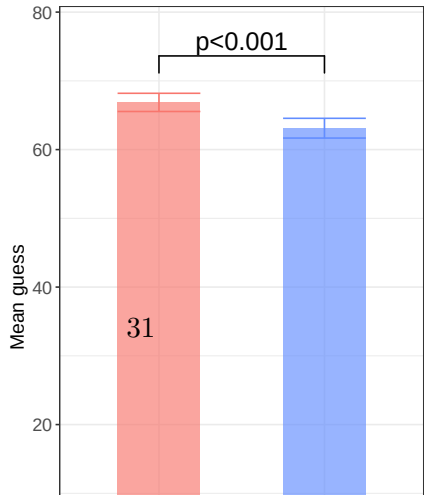
[Gullibility: word count]



[Bias: word count]



[Gullibility: scale]



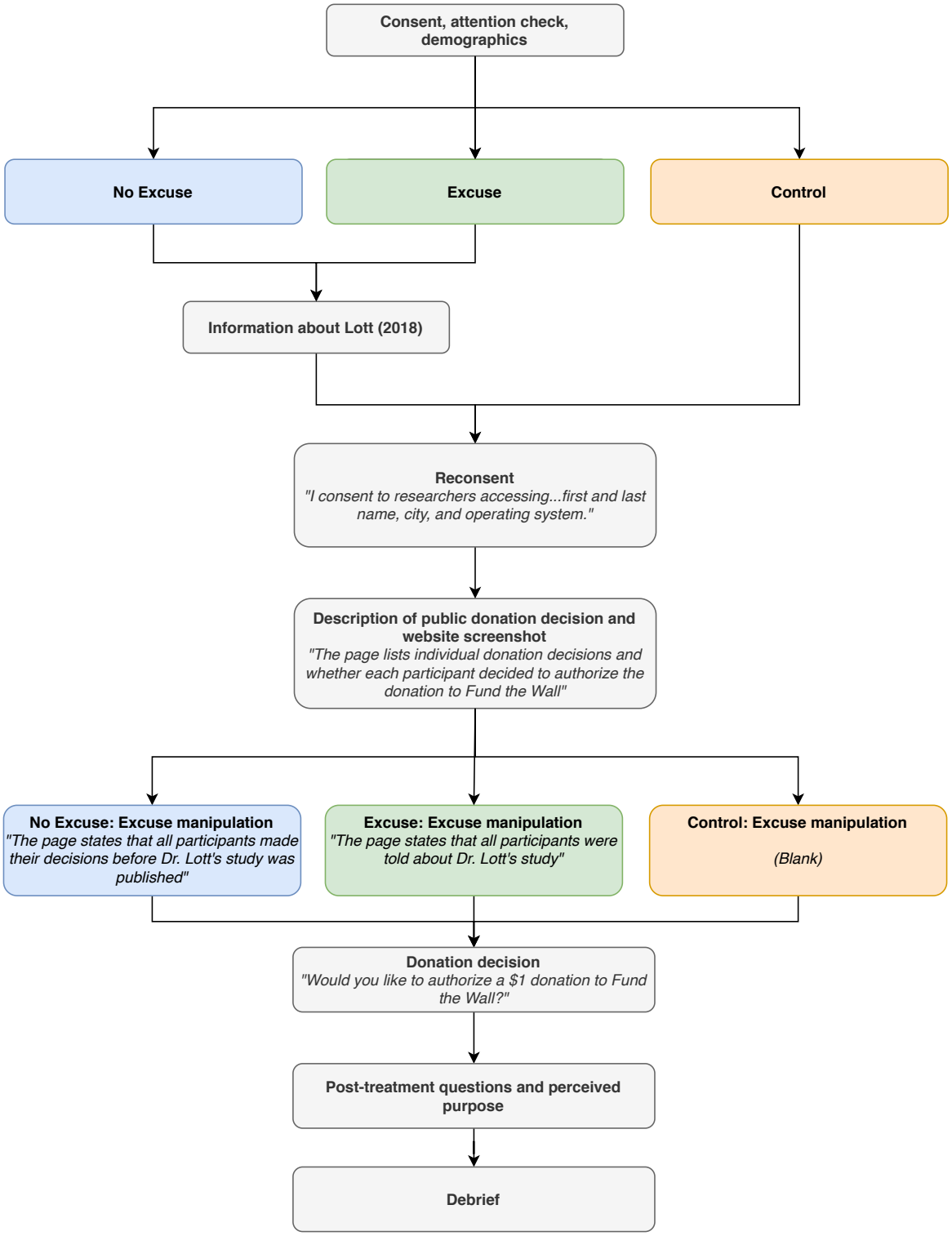


Figure 3: Experiment 2 – structure of design

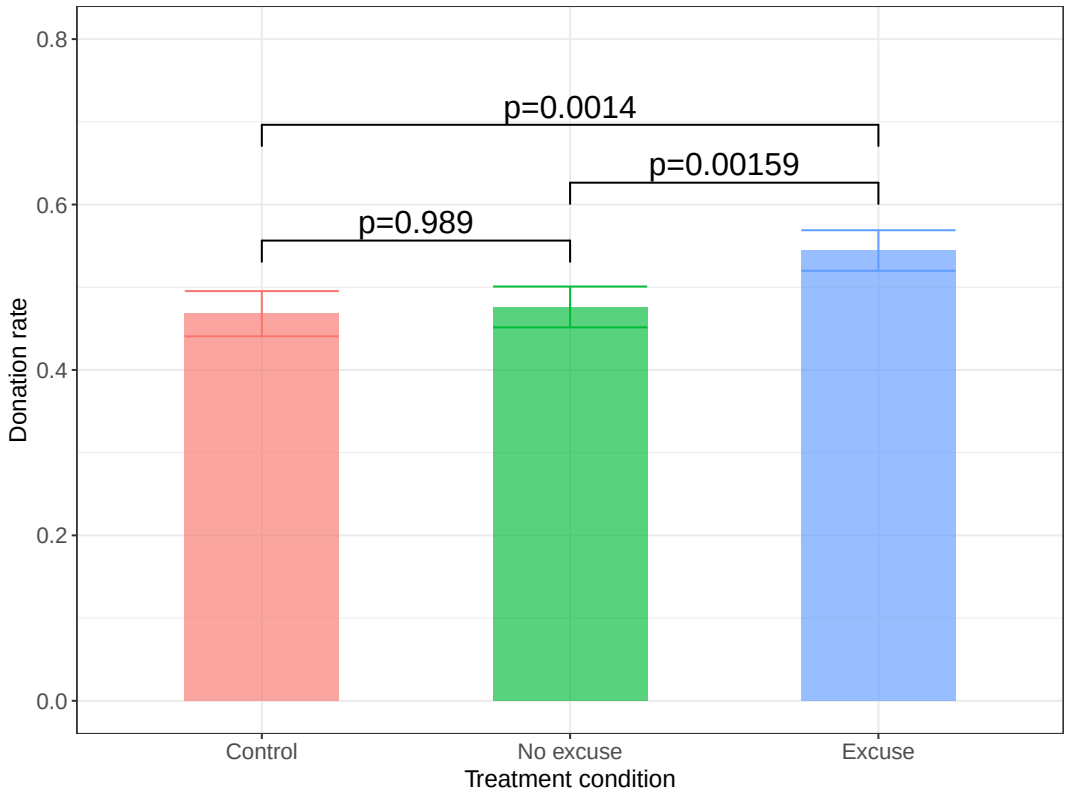


Figure 4: Experiment 2: Donation rates by group

Supplementary Appendix: not for publication

A Theory Proofs

A.1 Proof of Proposition 1

The tolerant sender ($i = 0$) chooses to donate ($a = 1$) if

$$\begin{aligned} v\mathbf{1}_{\{0=0\}} + \pi(0)\, b_0 + (1-\pi(0))\, b_1 = u_0(0, \pi(0)) < u_0(1, \pi(1)) = v\mathbf{1}_{\{1=0\}} + \pi(1)\, b_0 + (1-\pi(1))\, b_1 \\ \Leftrightarrow v + \pi(0)\, b_0 + (1-\pi(0))\, b_1 < \pi(1)\, b_0 + (1-\pi(1))\, b_1 \\ \Leftrightarrow v < (\pi(1) - \pi(0))(b_0 - b_1) \\ \Leftrightarrow \pi(1) - \pi(0) > \frac{v}{b_0 - b_1}, \end{aligned}$$

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

$$\begin{aligned} v\mathbf{1}_{\{0=1\}} + \pi(0)\, b_0 + (1-\pi(0))\, b_1 = u_0(0, \pi(0)) < u_0(1, \pi(1)) = v\mathbf{1}_{\{1=1\}} + \pi(1)\, b_0 + (1-\pi(1))\, b_1 \\ \Leftrightarrow \pi(0)\, b_0 + (1-\pi(0))\, b_1 < v + \pi(1)\, b_0 + (1-\pi(1))\, b_1 \\ \Leftrightarrow -v < (\pi(1) - \pi(0))(b_0 - b_1) \\ \Leftrightarrow \pi(1) - \pi(0) > -\frac{v}{b_0 - b_1}. \end{aligned}$$

A.2 Proof of Proposition 2

A.2.1 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-gullible sender to donate and the tolerant-gullible sender not to donate. If both the tolerant-nongullible and the intolerant-nongullible senders do not donate, 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-gullible and gullible 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-nongullible sender does not donate, the optimality condition for the intolerant-nongullible sender, (2), yields the second inequality.

$$\begin{aligned} a_0^* = 0 \Leftrightarrow & \pi^{NE}(1) - \pi^{NE}(0) = -\frac{p}{p + q(1-p)} \leq \frac{v}{b_0 - b_1}, \\ a_1^* = 0 \Leftrightarrow & -\frac{p}{p + q(1-p)} \leq -\frac{v}{b_0 - b_1} \\ \Leftrightarrow & \frac{p}{p + q(1-p)} \geq \frac{v}{b_0 - b_1} \\ \Leftrightarrow & v_1 \leq \frac{p(b_0 - b_1)}{p + q(1-p)}. \end{aligned}$$

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:

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

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

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

A.2.2 Excuse game

In the Excuse game, the receiver expects both types of gullible senders to donate. Since we look for an equilibrium in which the tolerant-nongullible sender does not donate and the intolerant-nongullible 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-nongullible sender does not donate, the intolerant-nongullible sender's optimality condition yields the first inequality:

$$\begin{aligned}
a_0^* = 0 &\Longrightarrow \pi^E(1) - \pi^E(0) = \frac{p(1-q)}{1-qp} - 1 \leq \frac{v}{b_0 - b_1} \\
&\Longrightarrow \frac{p(1-q) - 1 + qp}{1-qp} = -\frac{1-p}{1-qp} \leq \frac{v}{b_0 - b_1} \\
&\Longrightarrow -\frac{1-p}{1-qp} \leq 0 \leq \frac{v}{b_0 - b_1}, \\
a_1^* = 1 &\Longrightarrow \frac{p(1-q)}{1-qp} - 1 = -\frac{1-p}{1-qp} > -\frac{v}{b_0 - b_1} \\
&\Longrightarrow \frac{1-p}{1-qp} < \frac{v}{b_0 - b_1} \\
&\Longrightarrow v > \frac{(1-p)(b_0 - b_1)}{1-qp}.
\end{aligned}$$

We appeal to the intuitive criterion to rule out equliibria in which both tolerant- and intolerant-nongullible 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-nongullible 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-nongullible 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- nongullible 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-nongullible 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-nongullible 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-nongullible 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.

$$\begin{aligned}
0 &< \frac{p(b_0 - b_1)}{p + q(1-p)} - \frac{(1-p)(b_0 - b_1)}{1 - qp} \\
\implies \frac{1-p}{1-qp} &< \frac{p}{p + q(1-p)} \\
\implies (1-p)(p + q(1-p)) &< p(1-qp) \\
\implies p + q(1-p) - p^2 - pq(1-p) &< p - qp^2 \\
\implies q(1-p) - p^2 - pq + qp^2 &< -qp^2 \\
\implies q - p^2 - 2pq + 2qp^2 &< 0 \\
\implies q\left(1 - 2p + 2p^2\right) &< p^2 \\
\implies q &< \frac{p^2}{2p^2 - 2p + 1}.
\end{aligned}$$

B Appendix Figures and Tables

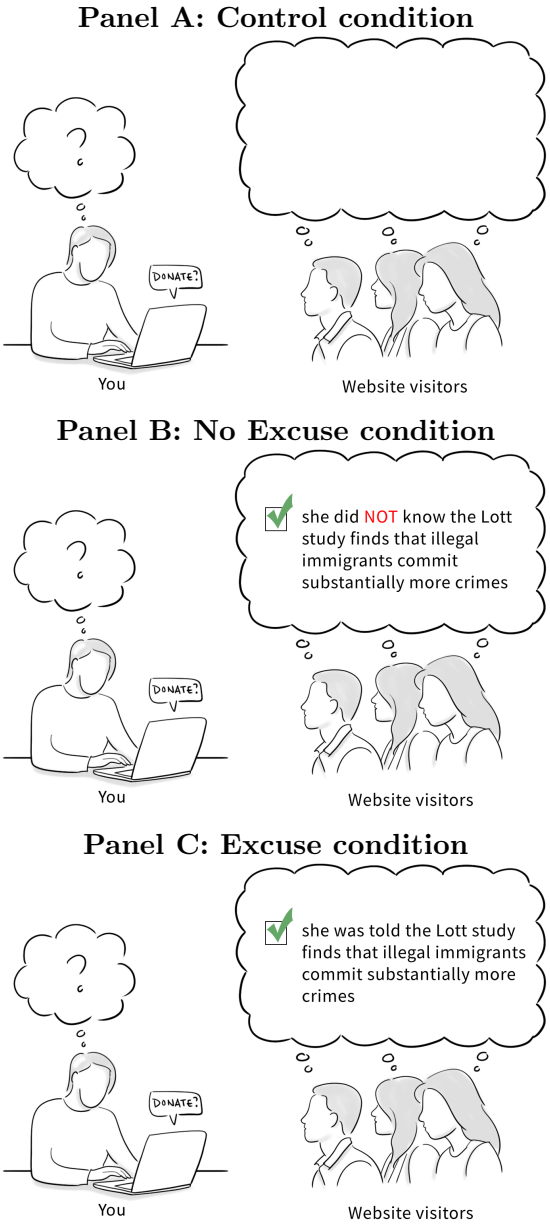


Figure A1: Illustrations of audience's information set

Table A1: Experiment 1: Sample representativeness

	Experiment 1	Pew
Variables:	(1)	(2)
Age	41.58	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	5151	4005

Notes: Mean of respondent characteristics in experiment 1 and the 2018 Pew Research Center's American Trends Panel Wave 39. Attriters dropped from sample.

Table A2: 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

[flushleft]

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

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

Dependent variable:
Attrited

Table A3: Experiment 1: Relationship between perceived motive and scores

	Dependent variable:					
	Inference about partner's score					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *Bias (z-score)*						
Used bias word	0.477***	0.473***	0.473***	0.502***	0.500***	0.500***
	(0.072)	(0.072)	(0.072)	(0.057)	(0.056)	(0.056)
Constant	−0.078***	−0.283	−0.301	−0.070***	−0.107	−0.149
	(0.027)	(0.278)	(0.281)	(0.021)	(0.220)	(0.222)
Observations	1,524	1,524	1,524	2,532	2,532	2,532
R^2	0.028	0.061	0.061	0.030	0.060	0.061
Adjusted R^2	0.027	0.053	0.052	0.030	0.056	0.056
Panel B: *Gullibility (z-score)*						
Used gullibility word	0.520***	0.486***	0.484***	0.429***	0.398***	0.397***
	(0.089)	(0.088)	(0.088)	(0.071)	(0.070)	(0.070)
Constant	−0.039	0.064	−0.055	−0.037*	−0.003	−0.107
	(0.026)	(0.300)	(0.302)	(0.021)	(0.232)	(0.234)
Observations	1,523	1,523	1,523	2,533	2,533	2,533
R^2	0.022	0.054	0.059	0.014	0.042	0.046
Adjusted R^2	0.021	0.046	0.051	0.014	0.037	0.041
Demographic controls	No	Yes	Yes	No	Yes	Yes
Partisan affiliation controls	No	No	Yes	No	No	Yes
Include pilot data	No	No	No	Yes	Yes	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. Columns 1-3 report results estimated on the sample from the main experiment, while Columns 4-6 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. The partisan affiliation control is an indicator that takes value 1 if the respondent self-reports being a strong Democrat, with weak Democrat as the reference category. Robust standard errors are reported.

(1)

Table A3 – *Continued from previous page*

	Attrited
	(1)
Age	0.001
	(0.002)
Age squared	0.00001
	(0.00002)
Black	−0.017
	(0.025)
Asian	−0.010
	(0.032)
White	−0.038*
	(0.022)
Hispanic	0.006
	(0.016)
Male	−0.029***
	(0.010)
High school	0.0004
	(0.039)
Some college, no degree	−0.023
	(0.038)
Associate degree	−0.030
	(0.040)
Bachelor degree	−0.034
	(0.038)
Post-bachelor degree	−0.053
	(0.040)
Strong Democrat	−0.013
	(0.011)
Excuse × Age	−0.001
	(0.002)
Excuse × Age squared	0.00001
	(0.00002)
Excuse × Black	0.046
	(0.034)
Excuse × Asian	−0.013

Continued on next page

	(0.044)
Excuse × White	0.045
	(0.030)
Excuse × Hispanic	0.030
	(0.023)
Excuse × Male	−0.016
	(0.015)
Excuse × High school	−0.008
	(0.048)
Excuse × Some college, no degree	−0.008
	(0.047)
Excuse × Associate degree	−0.023
	(0.050)
Excuse × Bachelor degree	−0.010
	(0.048)
Excuse × Post-bachelor degree	−0.010
	(0.051)
Excuse × Strong Democrat	−0.001
	(0.016)
DV mean (no excuse)	0.086
DV mean (excuse)	0.077
Observations	5,515
R^2	0.015
Adjusted R^2	0.010

Table A4: 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 A5: Experiment 2: Sample representativeness

	Experiment 2	Pew
	(1)	(2)
Panel A: *Republican*		
Age	47.11	49.50
Black	0.03	0.02
Asian	0.03	0.03
White	0.83	0.84
Hispanic	0.09	0.08
Male	0.49	0.51
Bachelors degree or higher	0.38	0.29
Observations	2022	2879
Panel B: *Independent*		
Age	43.53	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	2531	2622

Notes: Mean of respondent characteristics in experiment 2 and the 2018 Pew Research Center's American Trends Panel Wave 39. Attriters dropped from sample.

Table A6: 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	44.935	15.704	45.100	44.798	44.909	0.633	0.758	0.857
Black	0.076	0.266	0.069	0.088	0.072	0.085	0.797	0.135
Asian	0.043	0.202	0.041	0.042	0.045	0.877	0.657	0.772
White	0.821	0.383	0.826	0.815	0.823	0.465	0.877	0.556
Hispanic	0.110	0.313	0.113	0.107	0.111	0.644	0.854	0.776
Male	0.500	0.500	0.494	0.507	0.498	0.535	0.835	0.674
High school diploma	0.976	0.153	0.977	0.975	0.977	0.821	0.987	0.831
Bachelors degree	0.379	0.485	0.393	0.369	0.375	0.213	0.356	0.734
Republican	0.425	0.494	0.419	0.436	0.420	0.389	0.955	0.413

[flushleft]

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

Table A7: Experiment 2: Party heterogeneity

	Dependent variable:			
	Donated to Fund the Wall			
	Republicans		Independents	
	(1)	(2)	(3)	(4)
Excuse	0.068**	0.079***	0.068***	0.074***
	(0.028)	(0.024)	(0.024)	(0.023)
Control	−0.006	−0.008	0.003	0.013
	(0.028)	(0.026)	(0.024)	(0.024)
p-value (Excuse = Control)	0.0087	0.0011	0.0059	0.0089
Demographic controls	Yes	Yes	Yes	Yes
Partisan affiliation controls	Yes	Yes	Yes	Yes
Include pilot data	No	Yes	No	Yes
DV mean	0.670	0.674	0.356	0.358
DV std. dev.	0.470	0.469	0.479	0.469
Observations	1,582	1,961	2,146	2,471
R^2	0.071	0.049	0.133	0.049
Adjusted R^2	0.062	0.042	0.127	0.043

Notes: Columns 1-2 limit the sample to Republicans, while Columns 3-4 limit the sample to Independents. Columns 1 and 3 report results estimated on the sample from the main experiment, while Columns 2 and 4 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. Partisan affiliation controls include dummies for strong Republican, weak Republican, Republican-leaning Independent, and Democrat-leaning Independent. Robust standard errors are reported.

Table A8: Experiment 2: Perceived purpose of study

	Dependent variable:					
	Excuse (1)	Immigration attitudes (2)	Social image (3)	Information (4)	Persuasion (5)	Biased (6)
Excuse	−0.005	0.009	0.020**	0.012	−0.013	−0.00003
	(0.003)	(0.015)	(0.010)	(0.015)	(0.011)	(0.013)
Control	−0.003	0.133***	0.037***	−0.012	−0.081***	−0.036**
	(0.003)	(0.015)	(0.010)	(0.016)	(0.012)	(0.014)
p-value (Excuse = Control)	0.62	< 0.001	0.098	0.13	< 0.001	0.012
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Partisan affiliation controls	Yes	Yes	Yes	Yes	Yes	Yes
Include pilot data	Yes	Yes	Yes	Yes	Yes	Yes
DV mean	0.007	0.228	0.083	0.240	0.122	0.176
DV std. dev.	0.084	0.419	0.275	0.427	0.327	0.381
Observations	4,514	4,514	4,514	4,513	4,514	4,512
R^2	0.004	0.028	0.018	0.011	0.022	0.009
Adjusted R^2	0.001	0.025	0.015	0.007	0.019	0.006

Notes: The dependent variable in each column is an indicator for whether 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.

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. lc

Dependent variable:	
	Attrited
	(1)

Table A8 – *Continued from previous page*

	Attrited
	(1)
Age	−0.001
	(0.003)
Age squared	0.00004
	(0.00003)
Black	−0.008
	(0.048)
Asian	0.049
	(0.055)
White	0.036
	(0.040)
Hispanic	−0.027
	(0.030)
Male	−0.056***
	(0.017)
High school	−0.026
	(0.052)
Some college, no degree	−0.046
	(0.052)
Associate degree	−0.061
	(0.054)
Bachelor degree	−0.033
	(0.052)
Post-bachelor degree	−0.067
	(0.055)
Rep-leaning Ind	−0.004
	(0.023)
Weak Rep	−0.030
	(0.028)

Continued on next page

Strong Rep	-0.048^{**}
	(0.022)
Excuse × Age	-0.003
	(0.003)
Excuse × Age squared	0.00004
	(0.00003)
Excuse × Black	0.072
	(0.064)
Excuse × Asian	-0.027
	(0.077)
Excuse × White	0.007
	(0.051)
Excuse × Hispanic	0.019
	(0.040)
Excuse × Male	0.019
	(0.024)
Excuse × High school	0.035
	(0.068)
Excuse × Some college, no degree	0.007
	(0.068)
Excuse × Associate degree	0.051
	(0.072)
Excuse × Bachelor degree	0.062
	(0.069)
Excuse × Post-bachelor degree	0.052
	(0.073)
Excuse × Rep-leaning Ind	-0.004
	(0.032)
Excuse × Weak Rep	0.045
	(0.039)
Excuse × Strong Rep	0.021
	(0.031)
DV mean (no excuse)	0.151
DV mean (excuse)	0.159
Observations	3,792
R^2	0.031
Adjusted R^2	0.024

Table A9: 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 Survey instruments: Experiment 1

C.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.

>>

C.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**)

>>

C.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

>>

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**)

>>

C.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.

>>

D Survey instruments: Experiment 2

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

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.**

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

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

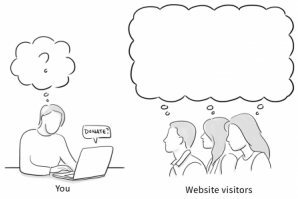
- **Alfred Marshall** was surveyed and **authorized** the donation to Fund the Wall
- **William Jevons** was surveyed and **did not authorize** the donation to Fund the Wall
- **Carl Menger** was surveyed and **did not authorize** the donation to Fund the

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

Reminder: what website visitors will learn

- They will learn whether or not you donated 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:

- 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?