Structural Analysis of Xenophobia

Huan Deng and Yujung Hwang *

April 10, 2023

Abstract

We estimate a signaling game of xenophobic behaviors to understand how individual racial animus and perceived unacceptance of racial animus determine xenophobic behaviors in equilibrium. To identify our model, we design a survey about anti-Chinese xenophobia in the US during the Pandemic. We validate our estimates by comparing our model predictions with the causal estimates obtained from an information Randomized Controlled Trial. We find raising perceived unacceptance is more effective than suppressing racial animus at reducing most xenophobic behaviors. We quantify the effects of a COVID infection on xenophobic behaviors in the short and long run.

JEL Classification: J15, Z13, Z18

Keywords: racial animus, perceived unacceptance, xenophobia, Sinophobia, COVID-19

1 Introduction

Many social interactions involve individuals displaying animosity towards different group members. One key feature of these interactions is that, while they contain a purely individual dimension – say, a personal dislike against individuals of a different race, religion, nationality, etc – they also have a social component: how others judge such animosity against a

*hdeng10@jhu.edu, yujungghwang@gmail.com. Our survey is registered at the AEA registry with ID AEARCTR-0006365 and it was approved by the Homewood IRB (Study Number: HIRB00011673). We thank Filipe Campante for reading our paper many times and giving insightful suggestions. The authors thank many people who kindly reviewed our survey questionnaire or draft and provided help, including Jane Chen, Taha Choukmane, Jonathan Elliott, Stelios Fourakis, Chao Fu, Matthew Hom, Louise Laage, Eunhee Lee, Lixiong Li, Tesary Lin, Mario Macis, Robert Moffitt, Aureo de Paula, Yin Pan, Zhenyu Wang, Pengpeng Xiao, Hanna Wang, Fangzhu Yang, and Matthew V. Zahn. We thank the participants at many seminars and conferences for their comments. Also, we thank the Hopkins Population Center and Hopkins Business and Health Initiative for funding this project. All errors are our own.

Electronic copy available at: https://ssrn.com/abstract=3958536

given group will affect one's propensity to express it. In particular, the latter can explain why such animosity can manifest quickly when people observe others' xenophobic behaviors. Understanding the interplay of these different dimensions is crucial for understanding the prevalence of such expression and figuring out how policy can address it.

This paper studies this interplay empirically in the context of rising xenophobia against Chinese immigrants in the US since the outbreak of the COVID-19 pandemic, which has been documented in several studies (Lu and Sheng (2022), Cao et al. (2022)). Our empirical analysis is grounded in theory to understand policy implications and long-run equilibrium outcomes. Using structural model estimates, we answer (i) which policy would be more effective in reducing xenophobic behaviors: should we reduce individual animosity or change perceptions about the reputational cost accompanied by xenophobic actions? (ii) how would the COVID-19 pandemic change xenophobia against Chinese immigrants in the short and long run? Previous literature lacks an explanation for these imperative questions. To illustrate, Paluck et al. (2021) conducted an extensive meta-analysis on 418 experiments concerning prejudice reduction interventions featured in 309 manuscripts between 2007 and 2019. Although they discovered that many interventions effectively altered discriminatory behaviors, they did not diminish negative stereotypes or animus. As a result, they raised the question, "What would this pattern imply theoretically?" (Paluck et al. (2021), p.553) We answer this question by providing a structural model to understand how animus and perceived reputational cost jointly determine the marginal change in xenophobic behaviors in equilibrium in the short and long run.

Our model is built upon Bénabou and Tirole (2006), which provided an important theoretical framework to study general pro-social behavior with an emphasis on the equilibrium nature of reputational motivation. In our model, an agent, characterized by own racial animus and perceived unacceptance of racial animus, decides whether to commit a xenophobic action. Two motivations underpin the xenophobic action decision. First, there is an intrinsic motivation – higher racial animus increases pleasure from a xenophobic action. Second, there is a reputational motivation – higher perceived unacceptability of racial animus leads to a higher perceived cost of a xenophobic action. Because racial animus is not observable to others, each agent uses a xenophobic action to signal own racial animus (type). The reputational response that an individual receives is called stigma/honor - that is, the expected racial animus conditional on xenophobic action/inaction. Stigma and honor are determined in equilibrium, reflecting which racial animus type commits a xenophobic action in the economy. Like Bénabou and Tirole (2006) model, our model may exhibit multiple equilibria: when people expect committing a xenophobic behavior would greatly harm one's reputation, then only a few people with very high racial animus would act xenophobically. When people expect the reputational cost is small, then even people with moderate racial animus may commit

Electronic copy available at: https://ssrn.com/abstract=3958536

xenophobic behaviors.

We carefully designed an online panel survey to measure key variables in our model while ensuring high-quality survey responses by adopting state-of-the-art conventions in the survey design literature. We collected 2,363 survey responses from non-Asians living in the US and stratified our sample by demographic characteristics. We model racial animus and perceived unacceptance as latent variables and identify them using multiple proxy variables for each (Cunha et al. (2010)). For racial animus, we used an established battery of questions from social psychology literature (Stephan et al. (1999))[1]. For perceived unacceptance, we developed a set of survey instruments, as we could not find existing survey questions for it.

Our measures for xenophobic behaviors include support for discriminatory institutions, outcomes from dictator games, and behaviors on Twitter. Support for discriminatory institutions is measured by hypothetical questions about whether to donate to a Sinophobic organization and whether to sign a Sinophobic petition. We argue that it is crucial to study these behaviors because such support can be contagious to ordinary people since they do not comprise a hate crime or violation of laws but still make it extremely difficult for Chinese immigrants to live in the US. Next, we implemented dictator games – money-splitting games – to measure altruism toward a Chinese immigrant relative to White Americans (Bertrand and Duflo (2017)). The games were incentivized with monetary compensation, with the maximum amount close to the base participation payment. If a respondent shares more money with a White American than with a Chinese immigrant, we code such behavior as xenophobic. Finally, we asked respondents to share a Twitter username if they have a Twitter account. We constructed variables on whether a respondent posted any pro-Asian or anti-Asian tweets during the pandemic. Most tweets were posted before participating in our survey, so these measures are least likely to be subject to the surveyor demand effect. Due to the small sample size and the selection in the merged Twitter data that we document later, we do not use Twitter-based measures for structural estimation. However, we use them to validate our survey instruments: whether someone posted any pro-Asian tweets is correlated with our survey instruments (measures on racial animus, perceived unacceptance, and xenophobic behavior).

Despite the presence of multiple equilibria, we show that the structural parameters can be point-identified (under some assumptions) when we observe multiple proxies for racial animus and perceived unacceptance. Multiple proxy variables are key to obtaining point identification: they identify the joint distribution of racial animus and perceived unacceptance, as well as the reputational gain that is specific to each equilibrium. And this identification is independent of any structural parameter. Once these objects are fixed, we are able to show the structural parameters are point-identified. One of the assumptions we need for identification

[1]The original questions are about Asian Americans. We replace the word Asian Americans with Chinese immigrants.

Electronic copy available at: https://ssrn.com/abstract=3958536

is that we know which observations were generated from the same equilibrium. We assume that our entire data was generated from the same equilibrium, for which we find supportive evidence. For estimation, we do not need a further assumption on equilibrium selection. In the counterfactual analysis, however, we need an equilibrium selection rule, and we select an equilibrium in which the reputational gain, defined as stigma minus honor, is closest to the baseline level.

To corroborate our structural estimation, we conducted an information Randomized Controlled Trial (RCT). The trial randomized individuals to either watch or not watch a 1-minute video [2] that aimed to influence Americans' perceptions of China and Chinese immigrants. We then compared our model predictions to the causal estimate of the effect of the information RCT. [3] The treated people who watched the video appeared to think xenophobia against Chinese immigrants is more socially acceptable, but we find little evidence that the video changes either racial animus or xenophobic behaviors.[4] Our model predictions on the Intention-to-Treat effects are close to the causal Intention-to-Treat estimates, which validates our model and the structural parameter estimates.

Using our estimated model, we present two main counterfactual analyses. First, we show which type of policies would be more effective in reducing xenophobia. We compare two types of policies, (i) policy raising perceived unacceptance (e.g. information intervention) and (ii) policy reducing racial animus (e.g. desegregation policy). We find raising perceived unacceptance is more effective than suppressing racial animus at reducing most xenophobic behaviors we consider.[5] To see this, we shift racial animus and perceived unacceptance distribution by the racial gap observed in our data, which is the difference between the most hostile racial group and the most friendly racial group[6] and we predict xenophobic behaviors using our structural parameter estimates. That is, we compare the counterfactual outcomes when the racial gaps in racial animus and perceived unacceptance disappear with the baseline outcomes. We find a much bigger decrease in most xenophobic behaviors when we shift the perceived unacceptance both in the short and long run.[7]

[2] You can watch the video at the following URL.
https://www.youtube.com/watch?v=8sjOWt6PWdA

[3] The power of our research design turned out to be lower than what we would have liked, so we do not use this variation for estimation.

[4] Statistically, we can not reject that the treatment effect on perceived unacceptance is the same as the treatment effect on racial animus.

[5] The only exception is the outcome from a dictator game, whose relative importance parameter for perceived unacceptance is estimated to be smallest, and for which reducing racial animus is marginally more effective than increasing perceived unacceptance.

[6] The most hostile racial group is white people and the most friendly group is the other race (non-white, non-black, and non-asian) people. The racial gap is 0.13 standard deviation each.

[7] In the short run, raising a perceived unacceptance leads to a decrease in xenophobic actions, ranging between -4% and -8%, whereas reducing a racial animus results in a decrease between -2% and -7%. In the long run, the differences are -6% to -16% decrease in actions when raising perceived unacceptance, and -2% to -9% decrease when reducing racial animus.

Electronic copy available at: https://ssrn.com/abstract=3958536

We propose two reasons why raising perceived unacceptance appears to be more effective than reducing racial animus in our counterfactual analysis. First, the marginal change in equilibrium depends on the mass of marginal agents, which in turn depends on the distributional shapes of racial animus and perceived unacceptance, as well as the current position of the indifference line to determine xenophobic behaviors. Our estimates imply that more marginal agents will opt out of xenophobic behaviors when perceived unacceptance shifts than when animosity shifts. This is the short-run effect after shifting each marginal distribution. Second, there is a long-run effect through changing reputational gain in equilibrium. Our estimates show that a much larger increase in reputational gain occurs when the distribution of perceived unacceptance is shifted to the right. Therefore, the reputational motivation causes marginal people with high perceived unacceptance to refrain from xenophobic behaviors.

The second counterfactual we consider is the effect of COVID infection and we find an optimistic result: COVID infection increases xenophobic behaviors in the short run, but such an increase is much milder in the long run, and in one case, it even decreases.[8] To make counterfactual predictions, we first estimate how COVID infection shifts the distribution of racial animus and perceived unacceptance using quantile regression with extensive controls, including proxies for pre-pandemic attitudes toward Chinese immigrants and the characteristics of social networks. Next, we predict using our structural parameter estimates how the equilibrium will change in the short run – defined as when reputational gain stays the same as a baseline – and in the long run – defined as when the reputational gain is updated to be consistent with new aggregate behaviors. We find the COVID infection polarizes the racial animus, shown by more mass at the tails, and this leads to an increase in xenophobic behaviors in the short run. However, in the long run, the increase is curved because the reputational gain from not making a xenophobic action increases as well. In a new equilibrium, xenophobic behaviors signal much higher racial animus because the pandemic increased the number of people with very high racial animus who newly engage in xenophobic behaviors. People with moderate racial animus then decide to quit xenophobic behaviors to avoid the additional stigma caused by the more extreme actors.

Our work extends the small literature on the structural estimation of Bénabou and Tirole (2006)-type model (Butera et al. (2022), Dubé et al. (2017)). Compared to Butera et al. (2022), we develop an empirical strategy to estimate any (pooling or separating) equilibrium of Bénabou and Tirole (2006)-type model, in which an action may not have a one-to-one mapping with (unobservable) types, and we allow for multidimensional types, including one that captures heterogeneous image concerns. Compared to Dubé et al. (2017), we propose an empirical strategy to achieve point identification despite potential multiple equilibria and to allow a flexible

[8] COVID infection is the only factor among various COVID-related experiences – including job loss – that significantly changes any motivation for xenophobic behaviors.

Electronic copy available at: https://ssrn.com/abstract=3958536

functional form of the marginal distribution of types. Aside from these papers, DellaVigna et al. (2016), and Karing (2019) estimate the value of social signaling without estimating the underlying structure, and our paper is differentiated from these papers in that we estimate the deep parameters behind the social signaling to simulate how the signaling would change in various counterfactual scenarios.

Finally, our structural work complements several reduced-form studies and applied theories on xenophobia. Lu and Sheng (2022) and Cao et al. (2022) documented the rise of xenophobia against Asians during the pandemic. We complement their findings by making a long-run prediction of the pandemic on xenophobia, which is infeasible without theory because the COVID-19 pandemic is a recent event at the time of writing this paper, and available data is not long enough to predict the long-run outcome. We find the pandemic's effect on anti-Chinese xenophobia can be different in the long run because of changing reputational gains associated with xenophobic (in)actions. Bursztyn et al. (2020) studied the effect of the rise of Donald Trump on the expression of xenophobic views and emphasized the role of perceived unacceptance on xenophobic behaviors. We strengthen this finding by providing a structural model to quantify the relative importance of racial animus and perceived unacceptance and to make a long-run prediction under various counterfactuals.

The remaining paper is structured as follows. Section 2 presents a signaling game of xenophobic behavior. Section 3 explains an identification and estimation strategy. Section 4 explains our survey design, and we relegate many details on the quality validation of our survey to Appendix Section B and Online Appendix Section F. Section 5 presents descriptive statistics and reduced-form evidence. Section 6 shows structural estimation results and the validation using the information RCT. Section 7 gives various counterfactual predictions. Section 8 concludes the paper.

2 A signaling game of xenophobic behavior

We adopt Bénabou and Tirole (2006)'s signaling game model and explain xenophobic behavior using two motivations: intrinsic motivation to express anti-Chinese animus and reputational motivation to maintain good social image[9].

[9] Bénabou and Tirole (2006) included extrinsic motivation in the model, but we omit this because, for most xenophobic behaviors we consider, extrinsic motivation, like a material payoff, is irrelevant.

Electronic copy available at: https://ssrn.com/abstract=3958536

2.1 An agent's problem

There is a continuum of agents whose types (ν, μ) are distributed according to a continuous joint distribution $F(\nu, \mu)$. ν is racial animus, and μ is the perceived social (un)acceptance of racial animus. Each agent chooses whether to commit a xenophobic action or not, $a \in \{0, 1\}$. When the agent chooses a xenophobic action, $a = 1$, the agent receives the utility gain that is equal to the racial animus ν and foresees the reputational return from the action, which is proportional to how other people would infer the agent's racial animus ν conditional on action $a = 1$, that is $E[\nu|a = 1]$. Each agent perceives differently how other people would tolerate the racial animus that is captured by μ. To sum up, the perceived stigma from a xenophobic action is $\mu E[\nu|a = 1]$. To match idiosyncratic dispersion in action, we allow idiosyncratic choice-specific Gumbel shock ϵ_1, ϵ_0. When the agent chooses not to commit the xenophobic action, $a = 0$, then the agent receives only the reputational gain, so-called perceived honor, $\mu E[\nu|a = 0]$. Both stigma and honor are determined at social equilibrium, reflecting who of which racial animus type ν commits a xenophobic action, and each agent is a small player who takes these reputational returns as given. So the agent's problem becomes the following:

$$\max_{a \in \{0,1\}} \quad (\nu - (\kappa\mu + c)E[\nu|a = 1] + \epsilon_1)a + (-(\kappa\mu + c)E[\nu|a = 0] + \epsilon_0)(1 - a) \tag{1}$$

$$(\nu, \mu) \sim F(\nu, \mu), \quad \epsilon_1, \epsilon_0 \overset{iid}{\sim} \text{Gumbel}(0, \beta), \quad \kappa > 0$$

κ is a scale parameter, and c is a location parameter for μ, and they jointly determine the relative importance of the image concern. β is a scale parameter for the Gumbel shocks ϵ_1, ϵ_0.

Our model reflects normalization choices. First, the location and scale of ν and the scale of μ do not affect the solution. However, the location of μ changes the counterfactual prediction[10]. Therefore, we add a location parameter for μ. The scale of the agent's problem is normalized by setting the coefficient in front of ν to be 1, and the κ and c capture the relative scale between ν and μ. The location and scale of ν, μ distribution, and $F(\nu, \mu)$, are later anchored using one proxy variable each (Assumption 2 (iv)). Note that after translating the μ distribution by $\frac{c}{\kappa}$ which can take any sign, the support of μ can include negative values - that is, we do not rule out the situation where racial animus is perceived as praiseworthy to some agents.

$$F(\nu, \mu) = C^{Joe}(F(\nu), F(\mu); \theta) \tag{2}$$

To model a joint density of the type (ν, μ), we model each marginal distribution and the dependence structure separately. Each marginal distribution can be fully nonparametric. The dependence structure is modeled with a Joe copula, which is an Archimedean copula with a

[10] We thank Chris Taber for this comment.

Electronic copy available at: https://ssrn.com/abstract=3958536

single parameter θ[11].

Next, we define an equilibrium in this signaling game.

Definition 1 (Equilibrium). *An equilibrium consists of an action $a^*(v,\mu,\epsilon_1,\epsilon_0)$ and the reputational gain $E^*[v|a=1]-E^*[v|a=0]$ such that*
1. For every individual, $a^*(v,\mu,\epsilon_1,\epsilon_0)$ is optimal given the reputational gain $E^*[v|a=1]-E^*[v|a=0]$. That is, $a^*(v,\mu,\epsilon_1,\epsilon_0)$ is a solution to the individual's problem defined in equation 1.
2. The reputational gain $E^*[v|a=1]-E^*[v|a=0]$ is consistent with the individual's behavior. That is,

$$E^*[v|a=1]-E^*[v|a=0]=\int_{(v,\mu,\epsilon_1,\epsilon_0)} v dF(v,\mu,\epsilon_1,\epsilon_0|a^*=1)-\int_{(v,\mu,\epsilon_1,\epsilon_0)} v dF(v,\mu,\epsilon_1,\epsilon_0|a^*=0) \tag{4}$$

As is well known, a signaling game may have multiple equilibria, and the conditions to have a unique equilibrium in a general signaling model are unknown.[12] We do not constrain our model to have a unique equilibrium but we prove that this does not cause an issue in identification (Proposition 1): one of the key assumptions we need for point identification is the assumption that we know which part of the data is generated from the same equilibrium.[13] Assumption 1 states that all the data is generated from the same equilibrium. One may be concerned if this assumption is violated because of geographic segregation or heterogeneous social networks. We examine whether reputational gains are substantially different across different US regions and neighborhoods displaying different political attitudes in our data and find they are not significantly different in both cases (Figure B.1, B.2 in Online Appendix). Therefore, we assume that the entire data is generated from the same equilibrium.

Assumption 1. *The data is generated from the same equilibrium.*

2.2 Measurement equations for proxies

We collect proxies for the types (v,μ) from our survey. Proxies are the noisy measurements of the types (v,μ), and we assume that we know the parametric relationship between proxies

[11]The copula choice was made after observing patterns in data. The Joe copula fits the empirical joint density well. The Joe copula formula is as follows :

$$C^{Joe}(u,v;\theta)=1-[(1-u)^\theta+(1-v)^\theta-(1-u)^\theta(1-v)^\theta]^{1/\theta},\quad \theta\in[1,\infty). \tag{3}$$

[12]Bénabou and Tirole (2006) provides conditions for a unique equilibrium in related but different models.
[13]For example, one can assume that the observations from the same group unit, such as a village or a school, are generated from the same equilibrium.

Electronic copy available at: https://ssrn.com/abstract=3958536

and the types summarized in the following equation:

$$Z_k^v = \alpha_{k0}^v + \alpha_{k1}^v v + \epsilon_k^v, \quad k \in \{1, \cdots, N_v\}, \quad \epsilon_k^v \overset{i.i.d.}{\sim} N(0, \sigma^2_{\epsilon_k^v}) \tag{5}$$

$$Z_g^\mu = \alpha_{g0}^\mu + \alpha_{g1}^\mu \mu + \epsilon_g^\mu, \quad g \in \{1, \cdots, N_\mu\}, \quad \epsilon_g^\mu \overset{i.i.d.}{\sim} N(0, \sigma^2_{\epsilon_g^\mu}) \tag{6}$$

We make the Assumption 2 regarding proxy variables and the latent types. Assumption 2 (i) assumes that the variances of latent variables are non-zero, so the types $\{v, \mu\}$ are heterogeneous. Assumption 2 (ii) states that there are multiple proxies, more than three each for (v, μ). This assumption is to identify measurement errors separately from the latent variables (Hu and Schennach (2008), Cunha et al. (2010)). Assumption 2 (iii) states that the factor loadings $\alpha_{k1}^v, \alpha_{g1}^\mu$ are not equal to zero, so proxies are relevant to learn about latent variables. Finally, Assumption 2 (iv) is to normalize the measurement equations for identification. The above system of equations is unidentified unless we normalize proxy variables[14]. Under Assumption 2 (iv), the location and the dispersion of the joint density of (v, μ), $F(v, \mu)$, are anchored using one proxy variable each for (v, μ), that is, $\{Z_1^v, Z_1^\mu\}$. In practice, how to choose the normalizing proxy variables matters. We discuss in Section 3 how we choose the normalizing proxy variables. Note that this normalization is innocuous because we allow for a location parameter c in the model to translate $F(v, \mu)$ along the μ dimension, which matters for counterfactual predictions. And the agent's problem in equation 1 implies that the location and scale of v, and scale of μ do not change the solution.

Assumption 2. *We make the following assumptions about the latent variables and their proxies.*
(i) (Non-Zero Variance) Assume $Var(v), Var(\mu) \neq 0$.
(ii) (Availability of Multiple Proxies) More than three proxies are available for each latent variable. That is, $N_v, N_\mu \geq 3$.
(iii) (Relevance) Assume $\alpha_{k1}^v, \alpha_{g1}^\mu \neq 0$ *for* $\forall k, g$.
(iv) (Anchorization/Normalization) Assume $\alpha_{10}^v = \alpha_{10}^\mu = 0$ *and* $\alpha_{11}^v = \alpha_{11}^\mu = 1$.

2.3 Auxiliary model for counterfactual analysis

We build an auxiliary model to explain how a factor D may shift the joint density of racial animus and perceived unacceptance, $F(v, \mu)$. Next, we make counterfactual predictions on how D changes an equilibrium on xenophobia using the structural model in equation 1.

[14]The exception is α_{10}^μ. An alternative identification strategy is to not anchor α_{10}^μ but omit c in equation 1. However, then it is difficult to use the sequential estimation strategy we use, which is key to gaining point identification in the presence of potential multiple equilibria. So we choose to anchor α_{10}^μ and allow for an additional location parameter c in equation 1.

Electronic copy available at: https://ssrn.com/abstract=3958536

For tractability, we assume the dependence between ν and μ stays invariant under counterfactuals, and a factor D shifts the marginal distribution of the key latent variables $\{\nu, \mu\}$. Admittedly, this assumption is strong, but previous literature also noted that this assumption keeps the estimation tasks feasible despite the curse of dimensionality. For example, Bayer et al. (2019) assumed that their marginal distributions could change, but the copula parameter for the joint density of state variables stays invariant over time due to computational challenges.

To estimate the marginal distributions of ν and μ, we use a set of quantile regressions in equations 7, and 8 where the dependent variables are the proxies of latent variables $\hat{\nu}, \hat{\mu}$, and the regressors are the factor of interest, D, and other covariates X. Later, we use the information RCT treatment, and the COVID-related experience as factors D shifting the distribution of $\{\nu, \mu\}$. The proxies of latent variables are constructed as the average of normalized proxies.[15] To interpret $\{\alpha^{\nu}(\tau), \alpha^{\mu}(\tau)\}$, the effect of a factor D as causal, we make either an independence assumption or conditional independence assumption - that is, the potential outcomes of (ν, μ) and the factor D are independent unconditionally or conditionally on covariates X[16]. Given the auxiliary model estimates and our structural parameter estimates $\{\kappa, c, \beta\}$, we can predict xenophobic behavior a under a counterfactual.

$$P\left[\hat{\nu} < D\alpha^{\nu}(\tau) + X\gamma^{\nu}(\tau) | D, X\right] = \tau \quad \text{a.s.} \tag{7}$$

$$P\left[\hat{\mu} < D\alpha^{\mu}(\tau) + X\gamma^{\mu}(\tau) | D, X\right] = \tau \quad \text{a.s.} \tag{8}$$

We define the short-run counterfactual outcome as the outcome when we hold the reputational gain fixed at the previous level. The long-run counterfactual outcome is defined as the outcome when we update the reputational gain to a new level consistent with the individual's behavior. It is a merit of a structural model to be able to produce long-run predictions even though the data covers a short time span.

Note that the long-run counterfactual outcome takes into consideration the social multiplier effect. The shift in the distribution of (ν, μ) will make the marginal types engage in or refrain from xenophobic behavior a. Next, the reputational gain $E[\nu|a = 1] - E[\nu|a = 0]$ will change reflecting the change in types who commit xenophobic behavior. And the change in reputational gain will make the marginal types change their xenophobic behavior. These updates will continue until the reputational gain becomes consistent with the individual's behavior.

[15] They are $\frac{\sum_k \tilde{Z}_k^{\nu}}{N^{\nu}}$, $\frac{\sum_g \tilde{Z}_g^{\mu}}{N^{\mu}}$ defined in equation 35, 36

[16] For information RCT treatment, an unconditional independence assumption is reasonable because we randomized the treatment. For COVID-related experiences, we rely on a conditional independence assumption.

Electronic copy available at: https://ssrn.com/abstract=3958536

3 Identification and Estimation

3.1 Identification

Despite multiple equilibria, we can achieve point identification under some assumptions when the proxies for the latent variables (ν, μ) – that is, $\{Z_k^\nu\}_{k=1}^{N_\nu}, \{Z_g^\mu\}_{g=1}^{N_\mu}$ – and action a are observed. We introduce the assumptions below. Assumption 3 is a standard one to conduct a deconvolution of densities (Cunha et al. (2010)). Assumption 4 rules out a corner solution $P(a=1)=0$ or $P(a=1)=1$. Later, we confirm that for all action measures from our survey, we have $0<P(a=1)<1$, so the Assumption 4 holds. Finally, Assumption 5 states that the coefficients in the logistic model (equation 9) are identified if we observe the true latent variables and the action, that is, $\{\nu, \mu, a\}$. This assumption rules out, for example, the joint density $F(\nu, \mu)$ that implies ν and μ are perfectly multicollinear. Again, we confirm that the estimated joint density $F(\nu, \mu)$ implies this assumption holds in our data.

Assumption 3 (Assumptions for Deconvolution). .
(i) Assume the characteristics function of a random vector of latent variables (ν, μ) is non-vanishing.
(ii) There exist two proxies for each latent variable (ν, μ) such that their normalized proxy vectors (W_1, W_2) guarantees the following expectation $E[iW_1 e^{i\zeta \cdot W_2}]$ exists for all $\zeta \in \mathbb{R}$.

$$
\begin{aligned}
W_1 &\equiv (\widetilde{Z}_k^\nu, \widetilde{Z}_g^\mu) = \left(\frac{Z_k^\nu - \alpha_{k0}^\nu}{\alpha_{k1}^\nu}, \frac{Z_g^\mu - \alpha_{g0}^\mu}{\alpha_{g1}^\mu} \right) \\
W_2 &\equiv (\widetilde{Z}_{k'}^\nu, \widetilde{Z}_{g'}^\mu) = \left(\frac{Z_{k'}^\nu - \alpha_{k'0}^\nu}{\alpha_{k'1}^\nu}, \frac{Z_{g'}^\mu - \alpha_{g'0}^\mu}{\alpha_{g'1}^\mu} \right), \quad k \neq k', g \neq g'
\end{aligned}
$$

Assumption 4 (Interior Solution). *Assume the action probability is strictly between 0 and 1,* $0<P(a=1)<1$.

Assumption 5 (Identification When Latent Variables Are Observed). *Suppose $\{\nu, \mu, a\}$ is observable. Then, the following coefficients, (ξ_0, ξ_1, ξ_2), in the logistic model (equation 9) are identified.*

$$
P(a=1|\nu,\mu) = \frac{exp(\xi_0 + \xi_1 \nu + \xi_2 \mu)}{exp(\xi_0 + \xi_1 \nu + \xi_2 \mu) + 1} \tag{9}
$$

And we state our point identification result in Proposition 1.

Proposition 1. *Suppose $\{a, \{Z_k^\nu\}_{k=1}^{N_\nu}, \{Z_g^\mu\}_{g=1}^{N_\mu}\}$ is observable. Under Assumption 1, 2, 3, 4, 5, the parameters in the measurement equations for proxies (equation 5, 6), the distribution $F(\nu, \mu)$, the reputational gain $E[\nu|a=1] - E[\nu|a=0]$, and the structural parameters (κ, c, β) can be uniquely identified.*

Electronic copy available at: https://ssrn.com/abstract=3958536

Proof. In the Appendix. □

The proof can be done in steps. First, given the data and Assumptions 1, 2, 3, 4, we can uniquely identify the parameters in the measurement equations for proxies (equation 5, 6), the distribution $F(\nu, \mu)$, and the reputational gain $E[\nu|a=1] - E[\nu|a=0]$. Next, given these objects identified and Assumption 5, we apply Theorem 1 in Hu and Ridder (2012) to show that the remaining structural parameters (κ, c, β) are point-identified.

3.2 Estimation

We estimate our model in several steps. Each step closely follows our identification proof for Proposition 1. First, we estimate the measurement equation parameters, the joint density $\widehat{F(\nu,\mu)}$, and the reputational gain $\widehat{E[\nu|a=1]} - \widehat{E[\nu|a=0]}$. Next, we estimate the structural parameter (κ, c, β) using the Indirect Inference given the other estimates.

We explain each estimation step below.

1. **Estimating measurement equation parameters**
   The measurement equation parameters $\{\alpha^{\nu}_{k0}, \alpha^{\nu}_{k1}, \alpha^{\mu}_{g0}, \alpha^{\mu}_{g1}, \sigma^2_{\epsilon^\nu}, \sigma^2_{\epsilon^\mu}\}$ can be estimated by replacing the moments in identifying equations to sample moments. The equations are suggested by Cunha et al. (2010).

$$Var(\nu) = \frac{\sum_{(k,k')} \frac{Cov(Z^{\nu}_1, Z^{\nu}_k) Cov(Z^{\nu}_1, Z^{\nu}_{k'})}{Cov(Z^{\nu}_k, Z^{\nu}_{k'})}}{\sum_{(k,k')} 1}, \quad 1 < k, k' < N^{\nu}, k \neq k' \tag{10}$$

$$Var(\mu) = \frac{\sum_{(g,g')} \frac{Cov(Z^{\mu}_1, Z^{\mu}_g) Cov(Z^{\mu}_1, Z^{\mu}_{g'})}{Cov(Z^{\mu}_g, Z^{\mu}_{g'})}}{\sum_{(g,g')} 1}, \quad 1 < g, g' < N^{\mu}, g \neq g' \tag{11}$$

$$E[\nu] = E[Z^{\nu}_1] \tag{12}$$

$$E[\mu] = E[Z^{\mu}_1] \tag{13}$$

$$\alpha^{\nu}_{k1} = \frac{Cov(Z^{\nu}_1, Z^{\nu}_k)}{Var(\nu)} \tag{14}$$

$$\alpha^{\mu}_{g1} = \frac{Cov(Z^{\mu}_1, Z^{\mu}_g)}{Var(\mu)} \tag{15}$$

$$\alpha^{\nu}_{k0} = E[Z^{\nu}_k] - \alpha^{\nu}_{k1} E[\nu] \tag{16}$$

$$\alpha^{\mu}_{g0} = E[Z^{\mu}_g] - \alpha^{\mu}_{g1} E[\mu] \tag{17}$$

$$\sigma^2_{\epsilon^\nu} = Var(Z^{\nu}_k) - (\alpha^{\nu}_{k1})^2 Var(\nu) \tag{18}$$

$$\sigma^2_{\epsilon^\mu} = Var(Z^{\mu}_g) - (\alpha^{\mu}_{g1})^2 Var(\mu) \tag{19}$$

Electronic copy available at: https://ssrn.com/abstract=3958536

In practice, the choice of normalizing proxy variables $\{Z_1^\nu, Z_1^\mu\}$ matters. We choose proxy variables $\{Z_1^\nu, Z_1^\mu\}$, that are most correlated with other proxy variables to anchor the location and the scale of (ν, μ). This is to reduce the variance of measurement equation parameter estimates, as the covariance between the anchor proxy variable and another proxy variable inversely affects the factor loading parameter through changing $Var(\nu)$ (see equation 15). Therefore, large covariance between the anchor proxy variable and other proxy variables helps reduce the variance in the estimates. We replicated our main results when we use different proxy variables to anchor the location and the scale of (ν, μ). Structural parameter estimates change because of different normalizations. We confirmed most model predictions remain qualitatively similar.[17]
Note that $Var(\nu), Var(\mu)$ are overidentified in the model because $Var(\nu) = \frac{Cov(Z_1^\nu, Z_k^\nu) Cov(Z_1^\nu, Z_{k'}^\nu)}{Cov(Z_k^\nu, Z_{k'}^\nu)}$ for $\forall k, k'$ such that $k, k' \neq 1, k \neq k'$ and $Var(\mu) = \frac{Cov(Z_1^\mu, Z_g^\mu) Cov(Z_1^\mu, Z_{g'}^\mu)}{Cov(Z_g^\mu, Z_{g'}^\mu)}$ for $\forall g, g'$ such that $g, g' \neq 1, g \neq g'$. We take an average of results obtained from all possible pairs $(k, k'), (g, g')$ to estimate $Var(\nu)$ and $Var(\mu)$.

2. **Estimating the marginal densities of ν and μ**
Using the estimates of the measurement equation parameters, we construct normalized proxies, which have error-in-variable structures.

$$\widetilde{Z}_k^\nu = \frac{Z_k^\nu - \alpha_{k0}^\nu}{\alpha_{k1}^\nu} = \nu + \widetilde{\epsilon}_k^\nu, \quad k \in \{1, \cdots, N_\nu\}, \quad \widetilde{\epsilon}_k^\nu \overset{i.i.d.}{\sim} N\left(0, \left(\frac{\sigma_{\epsilon^\nu}}{\alpha_{k1}^\nu}\right)^2\right) \tag{20}$$

$$\widetilde{Z}_g^\mu = \frac{Z_g^\mu - \alpha_{g0}^\mu}{\alpha_{g1}^\mu} = \mu + \widetilde{\epsilon}_g^\mu, \quad g \in \{1, \cdots, N_\mu\}, \quad \widetilde{\epsilon}_g^\mu \overset{i.i.d.}{\sim} N\left(0, \left(\frac{\sigma_{\epsilon^\mu}}{\alpha_{g1}^\mu}\right)^2\right) \tag{21}$$

Next, we apply a Li and Vuong (1998) deconvolution kernel estimator to the normalized proxy variables to estimate the nonparametric density of ν and μ. We follow Delaigle and Gijbels (2004) and Kato et al. (2021) to choose the bandwidth for Li and Vuong (1998) estimator.[18]

3. **Estimating the Joe copula parameter θ**
We estimate the Joe copula parameter θ to match the correlation between average z-

[17] This result is available upon request.

[18] The bandwidth parameter was chosen to minimize the AMISE(h) formula presented in the Appendix B.2 of Kato et al. (2021); The kernel function used in Li and Vuong (1998) estimator is a flat-top kernel $\phi_K(s)$ (see below equation) with tuning parameters b=1, c=0.05.

$$\phi_K(s) = \begin{cases} 1 & \text{if} \quad |s| \leq c \\ exp(-bexp(-b/(|s|-c)^2)/(|s|-1)^2) & \text{if} \quad c < |s| < 1 \\ 0 & \text{if} \quad |s| \geq 1 \end{cases} \tag{22}$$

Electronic copy available at: https://ssrn.com/abstract=3958536

scores of proxies of racial animus and perceived unacceptance with that of a simulated sample; To simulate the average z-scores of proxies of racial animus and perceived unacceptance, we first draw $(F(\nu), F(\mu))$ assuming a Joe copula with the parameter θ and apply the inverse of the marginal CDFs of ν and μ that are estimated in Step 2. This gives a draw of (ν, μ). Next, we simulate the proxies $\{Z_k^\nu\}, \{Z_g^\mu\}$ given the (ν, μ) draw using the estimated parameters in Step 1. The simulation sample size is five times larger than our data size.

4. **Estimating the reputational gain** $\widehat{E[\nu|a=1]} - \widehat{E[\nu|a=0]}$
We estimate the reputational gain using the normalized proxy variables of racial animus.

$$\widehat{E[\nu|a=1]} - \widehat{E[\nu|a=0]} = \frac{\sum_k \sum_i \tilde{Z}_{ik}^\nu \mathbb{1}(a_i = 1)}{\sum_k \sum_i \mathbb{1}(a_i = 1)} - \frac{\sum_k \sum_i \tilde{Z}_{ik}^\nu \mathbb{1}(a_i = 0)}{\sum_k \sum_i \mathbb{1}(a_i = 0)} \tag{23}$$

5. **Estimating the structural parameters** (κ, c, β) **through Indirect Inference**
We estimate the structural parameters (κ, c, β) by Indirect Inference (Gourieroux et al. (1993)).[19] We let the structural parameters vary by xenophobic actions. We have three xenophobic action measures, so we estimate nine structural parameters in total. For each xenophobic action, we match five moments: regression coefficients $\{\xi_0, \xi_1, \xi_2\}$ regressing the xenophobic action on average z-score of racial animus and perceived unacceptance, average xenophobic action $P(a=1)$, and the model predicted reputational gain $E[\nu|a=1] - E[\nu|a=0]$.

$$P(a=1|\{\tilde{Z}_k^\nu\}, \{\tilde{Z}_g^\mu\}) = \xi_0 + \xi_1 \left(\frac{\sum_k \tilde{Z}_k^\nu}{N^\nu} \right) + \xi_2 \left(\frac{\sum_g \tilde{Z}_g^\mu}{N^\mu} \right) \tag{24}$$

We use a diagonal weighting matrix with a diagonal that includes the inverse of the variance of each moment (Altonji and Segal (1996)). We estimate the variance of each moment using 100 bootstrap samples.
The objective function of Indirect Inference is non-differentiable due to discreteness in a choice variable. To smooth the objective function, we use a simulation sample five times larger than our data size, and we do an extensive grid search and use the Nelder-Mead algorithm (Nelder and Mead (1965)) for estimation.[20]

To account for the cumulation of sampling errors, the standard errors are computed by re-

[19] That is, given the estimates from previous steps and structural parameters (κ, c, β), we can simulate observations $\{a, \{\tilde{Z}_k^\nu\}, \{\tilde{Z}_g^\mu\}\}$ and assess the model fit by comparing moments in data and simulated data. The parameter estimates for (κ, c, β) are the ones that minimize the weighted distance between data moments and simulated data moments.

[20] We have considered using a generalized Indirect Inference (Bruins et al. (2018)) but decided not to use it because it requires a substantial amount of smoothing to remove kinks in our data but then it brings too much bias in the estimates.

Electronic copy available at: https://ssrn.com/abstract=3958536

peating the entire estimation procedure 100 times using a bootstrap sample with replacement.

4 Survey Instruments

For brevity, we defer much of the discussion on survey design regarding survey quality and potential social desirability bias to Appendix B, and we keep the minimal discussion on our key survey instruments in this section.

We conducted a 15-minute online survey through a survey firm Respondi. The firm Respondi sent invitation emails to their panel members. The survey started on March 24, 2021, and finished on May 24, 2021. After dropping low-quality responses, our sample includes 2,363 non-Asian individuals living in the US, who are aged between 18 and 70 years old. We stratified our sample in terms of gender, race, education, age, marital status, and income[21]. We paid $2.25 for each complete 15-minute survey and we paid extra rewards based on their answers. Respondi survey firm has set different compensation amounts to the survey participants based on their demographics.

We used the consistent wording "Chinese immigrants (living in the US)" throughout our survey, not to confuse it with either Chinese people living abroad or the Chinese government. Online Appendix C shows a few selected survey questions and provides a link to take our survey online. The complete survey questionnaire can be found on the author's website.[22] Appendix Figure B.1 shows our survey flow.

4.1 Measurement of Anti-Chinese Racial Animus and Perceived Unacceptance of Racial Animus

This section explains our survey instruments to measure anti-Chinese racial animus and perceived unacceptance of racial animus. Table 1 lists our survey instruments.

To measure anti-Chinese racial animus, we used a subset of questions developed by Social

[21] We exclude a non-Chinese Asian sample from the analysis. Non-Chinese Asians comprise only 3% of the US population, and therefore, including this population in the analysis will not change our results much. Understanding non-Chinese Asians' Sinophobia may be interesting, but this will require oversampling non-Chinese Asians, which was not feasible in our project due to cost concerns. Another reason why they are excluded is that non-Chinese Asians may have very different motivations for their bias against Chinese people, so including them would make our sample more heterogeneous. Many non-Chinese (east) Asians are difficult to distinguish from Chinese individuals physically, so they also became victims of hate crimes during the pandemic (Tessler et al. (2020)).

[22] https://github.com/yujunghwang/YujungHwangPaper/blob/main/StructuralXenophobia/SurveyTreatedGroup.pdf

Electronic copy available at: https://ssrn.com/abstract=3958536

Table 1: Survey instruments for anti-Chinese racial animus and perceived unacceptance of racial animus

Anti-Chinese Racial Animus	
Question	Please indicate your agreement with each of the following statements concerning immigration from China. (Slider scale from 0(strongly disagree) to 100(strongly agree))
Statement	• The children of Chinese immigrants should have the same right to attend public schools in the U.S. as Americans do. • Chinese immigrants do **not** pose any health hazards to U.S. citizens.
Question	For each of the items listed below indicate what your attitudes are toward Chinese immigrants. My attitude toward Chinese immigrants is : (Slider scale from 0(strongly disagree) to 100(strongly agree))
Statement	• Warmth • Hatred • Acceptance • Dislike
Perceived Unacceptance of Racial Animus	
Question	Please indicate your agreement with each of the following statements concerning the attitudes of **typical American citizens (below referred to as "people").** (Slider scale from 0(strongly disagree) to 100(strongly agree))
Statement	• Even though someone publicly says he/she dislikes Chinese immigrants, the person can still maintain good social relationships with most people. • People will criticize if someone calls the novel coronavirus "China virus." • It is **not** shameful to publicly support reducing the number of Chinese immigrants because they threaten our prosperity. • It is **not** socially acceptable to make a public statement that the US government should restrict the rights of Chinese immigrants.

Psychologists (Stephan et al. (1999))[23]. We did not use their entire survey to limit the length of ours. The questions consisted of two types of questions. The first type of questions asked how much respondents agree with each statement about Chinese immigrants in the US, which can reveal racial animus against Chinese immigrants. The second type of questions inquired about the feelings towards Chinese immigrants. These questions are jointly used to identify a single latent variable, called 'anti-Chinese racial animus' v in the model.

Unlike the measures for anti-Chinese racial animus, we could not find similar survey instruments for the perceived unacceptance of racial animus against Chinese or Asian immigrants. Therefore, we developed our own survey instruments. The statements described how the *typical American citizens* would react to or judge a Sinophobic behavior made in public. These statements are distinguished from the statements about racial animus in that they are

[23]We took a subset of questions developed by Walter G. Stephan. http://psych.nmsu.edu/faculty/walter/asian_questionaire.pdf

Electronic copy available at: https://ssrn.com/abstract=3958536

about the reactions of typical American citizens, not the survey respondents. The behaviors included in the statement are intuitively Sinophobic: the behaviors are publicly announcing dislike against Chinese immigrants, calling the novel coronavirus a “China Virus", and publicly supporting reducing the number of Chinese immigrants on the ground that they threaten the prosperity of the US, publicly claiming to restrict the rights of Chinese immigrants.

The fact that we are using multiple statements for each construct, instead of only one statement, alleviates any potential concern one may have on any single statement alone. We pool information from multiple responses to related questions to estimate the distribution of latent variables ν, μ. Moreover, our measurement equations 5, 6 account for the fact that some statements might be weakly related to the latent variable of interest than others, might be subject to bigger measurement errors, or can have a very different average response which can be represented as a different location parameter $\alpha^{\nu}_{k0}, \alpha^{\mu}_{g0}$ in equation 5, 6.

We have strong evidence that the responses to these survey instruments are not cheap talk. First, we show that these survey instruments have high internal consistency in our Online Appendix (Table A.1 and A.2). This is consistent with our assumption that the statements of each group measure the same latent variable. Second, both proxies for racial animus and perceived unacceptance strongly and significantly predict xenophobic behaviors (Table 4). If the responses to the measures of racial animus or perceived unacceptance were cheap talk, for example, due to tremendous social desirability bias, they must not predict xenophobic behaviors, which is not the case in our sample.

There is little concern about social desirability bias in the responses to these questions. First, we implemented List randomization for these questions and did not find evidence of social desirability bias. See Appendix Section B for details. Second, by estimating our measurement equation 5, 6, we remove any social desirability bias in the form of shifting the location of the responses or changing the dispersion of the responses under the parametric assumption we make. Finally, our survey instruments on racial animus have been established in the literature (Stephan et al. (1999)), and the feeling thermometer questions are widely used in many social surveys, including American National Election Studies (ANES) and World Value Surveys (WVS).

4.2 Measurement of Xenophobic Behavior

We collected a wide range of hypothetical and incentivized xenophobic behavior measures for a complete picture, including whether to donate to a Sinophobic institution, whether to sign a Sinophobic petition, and dictator game outcomes to measure altruism toward Chinese immigrants relative to white Americans. We also collected tweets made by survey participants

Electronic copy available at: https://ssrn.com/abstract=3958536

during the pandemic by asking for their Twitter usernames and used them to validate our survey instruments. To keep the discussion short, we defer the discussion on using Twitter-based measures for validation to the Online Appendix Section F.

The donation and petition questions were hypothetical choice questions.[24] One shortcoming of using hypothetical choice questions is that respondents may not consider their choices seriously. To complement these questions, we made our participants play a dictator game, which was incentivized with real money at stake, and we also measured real-world behaviors on Twitter from a small subsample. Most tweets were made before participating in our survey, so they are least likely to be subject to surveyor demand effect.[25] However, we do not use Twitter data for structural estimation because of the small sample size and selection in Twitter data.

Below, we briefly describe each of these behavioral measures. You can find the questions we used in Online Appendix C.4. First, in the donation question, we gave a short description of two different organizations with opposing stances on Chinese immigrants. One organization defined Chinese students and scholars as potential spies and urged restricting the entry of Chinese students and scholars into the US. The other organization made the opposite claim. We asked if respondents would like to donate $1 hypothetically to either organization. If respondents chose an organization with a hostile attitude toward Chinese students and scholars, we coded it as xenophobic behavior.

Second, we gave two short petitions for participants to review. One petition called for national efforts to protect US security and wealth from the threats posed by Chinese immigrants. The other petition urged defending the Chinese immigrants' safety and rights. If respondents opted to sign the former petition, we coded it as xenophobic behavior.[26]

Third, our dictator game was incentivized with real money at stake and did not include any deception. We recruited one Chinese immigrant and one white American to become a receiver player and paid them according to the dictator game outcomes. Every survey participant played the dictator game twice with both receiver players in random order. We showed the receiver players' headshot photos and their first names, which signal their ethnicity, and asked them to choose how much to share with the receiver players if they were given $1 to share. If participants shared more money with a White American than with a Chinese immigrant, we coded it as xenophobic behavior.[27] $1 is small but is close to the base participation

[24] We could not implement *real* donation and *real* petition questions because of an objection from our Institutional Review Board, even though such designs were used and published by others (Bursztyn et al. (2020), Grigorieff et al. (2018), Elías et al. (2019)).

[25] This is under the assumption that those who shared their Twitter usernames did not selectively delete their previous tweets after participating in our survey. It is not possible to verify this assumption, unfortunately.

[26] We allow participants not to sign any petition if they want. 26% of the sample did not sign any petition.

[27] In our reduced-form analysis, we present the results using the share difference between a White American and a Chinese immigrant. We do not include the share difference in our structural analysis because our model explains a discrete binary choice.

Electronic copy available at: https://ssrn.com/abstract=3958536

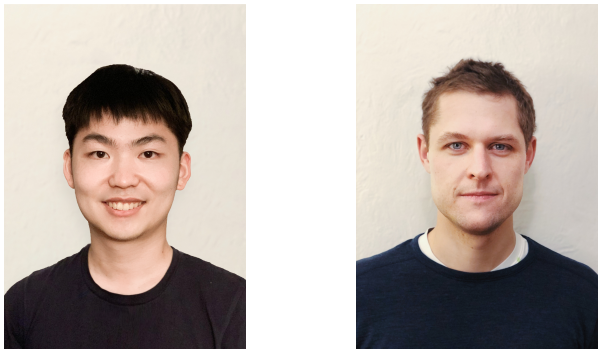


Figure 1: Pictures of the receiver players in a dictator game.
Left player's name: Haozheng, Right player's name: Peter

reward. Therefore, by making the most selfish decisions, participants can earn up to 200% of the base participation reward if they are selected to be paid.[28] To save on survey costs, we told them that 10% of our sample would be randomly selected to be paid according to their responses. We emphasized that their choices in the game would not affect the probability of being selected for payment.

Admittedly, our xenophobic action measures other than Twitter measures are not observed by the public but only by research team members. In that sense, one may question why these behaviors may rely on the image concern. We argue that our xenophobic action measures still reflect the reputational concern, and we offer three explanations for this. First, donation and petition questions are generally considered social actions and our hypothetical question can be understood as asking what the respondent would do if they can make the same decision in real life. In this consideration, participants can factor in the potential reputational concern following the decisions, if they were assuming a hypothetical real-world situation where these actions can be observed by the public. The fact that our perceived unacceptance measures can explain the xenophobic behavior measures is consistent with this explanation that survey participants factored in the potential reputational concern following such decisions. Second, even in the absence of publicity, people may judge themselves through the lens of the public. For example, Dubé et al. (2017) has shown that people consider image concern even in the absence of publicity of their donation actions, which can be explained by self-signaling. Although the publicity condition was ambiguous in our questions, participants may have judged their actions in the eyes of the public. We conjecture that increasing the publicity of action may still affect the average behavior, but exploring this is beyond the scope of our study. Finally, the sizable correlation between posting any pro-Asian tweets in the real world with other xenophobic action measures in our survey reassures that these xenophobic action measures are not likely to be cheap talks and reflect the reputational concern (Online Appendix

[28] This is because they play the dictator game twice with a different receiver player.

Electronic copy available at: https://ssrn.com/abstract=3958536

Section F).

4.3 Information Randomized Controlled Trial

We showed a one-minute video[29] about how the pandemic is changing the perceptions of China and Chinese immigrants to a randomly chosen half of the participants. The purpose of this video is to shift the distribution of μ, the perceived unacceptance of racial animus. The video reports the research findings from a nonpartisan think tank, and it includes no deception. We incentivized viewers to pay more attention to the video.[30]

We distinguish the group randomized into treatment from the group that got effectively treated. We consider a participant was 'effectively treated' if they answered a post-treatment question about the video content correctly and if they reported no technical issue in playing the video afterward. Later, we present both intention-to-treat (ITT) estimates and local average treatment effect (LATE) estimates. To compute the local average treatment effect, we instrument the effective treatment using randomization into a treatment group. The complier rate was high, 87%.

We measure racial animus and perceived unacceptance after the information RCT (see Appendix Figure B.1 for a survey flow). Therefore, we can not estimate the heterogeneous treatment effect by prior attitude and belief. However, this does not affect our structural estimation results because we have shown that the structural parameters can be point-identified without the variation from the information RCT in Proposition 1. Also, eliciting both prior and posterior attitudes and beliefs comes at a cost: doing so would increase the survey length that can potentially increase attrition, could potentially worsen the surveyor demand effect by asking the same questions twice, and could introduce the consistency bias when respondents answer the same questions the second time. For further discussion, see Haaland et al. (2022).

We find that the information treatment does not change social desirability bias in responses to questions about racial animus and perceived unacceptance of racial animus. This is shown in Figure D.1 in Online Appendix Section D. Figure D.1 compares the List randomization reports with direct reports by the treatment status. For both treated and control groups, the means from the List randomization are not statistically different from the means from those of the direct report. Therefore, differences in the racial animus or perceived acceptance proxies by treatment status reflect the change in the latent variables of racial animus

[29] You can find the video we used from the YouTube link.
https://www.youtube.com/watch?v=8sjOWt6PWdA

[30] We told respondents before the treatment, they would be given a lottery to win a small reward, worth the same as the base participation payment, if they answer correctly about the video content later. Afterward, we asked whether they had any technical issues playing the video on their device.

Electronic copy available at: https://ssrn.com/abstract=3958536

or perceived unacceptance, and these are not the artifact of different measurement errors by treatment status.

We find no evidence that the information treatment increases sample attrition. Table D.1 in Online Appendix shows that the RCT treatment increases neither the passage rate of the second attention check nor whether stating any bias in our survey. The fact that the RCT treatment does not affect the surveyor demand effect reassures that there is no evidence of an increase in social desirability bias for the treated group.

5 Descriptive Statistics and Reduced-Form Evidence

This section presents descriptive statistics about our survey sample and the reduced-form evidence which supports our model.

5.1 Descriptive Statistics

Our sample matches the non-asian US population reasonably well, although not perfect due to the limitations of an online panel survey. We re-weight our sample to match the non-asian US population in our reduced-form analysis using the weight provided by the survey firm. Table 2 compares our sample with the characteristics of the representative non-asian US population. Our sample matches this population well in terms of gender, race, and marital status. We have fewer young people aged between 18 and 29, more older people aged between 60 and 70, more lower-income people, fewer higher-income people, fewer people from the West, and more people from the Northeast and Midwest.

We found a substantial racial gap in racial animus and perceived unacceptance (Figure 2), and later we will show how the xenophobia equilibrium might change if the racial gaps in racial animus and perceived unacceptance disappear respectively. White respondents have notably higher racial animus and lower perceived unacceptance compared to other groups. Whiskers show that these differences are statistically distinguishable from either black respondents or other race. Black respondents' racial animus is not significantly lower than that of white respondents, but they show the highest perceived unacceptance of racial animus.

We asked about COVID-related experiences to understand how xenophobia could have been different if a COVID outbreak had never occurred. Table 3 shows the summary statistics of COVID-related experience in our sample. A substantial number of respondents either got infected with COVID or know someone close to them who did; specifically, about 8% of our

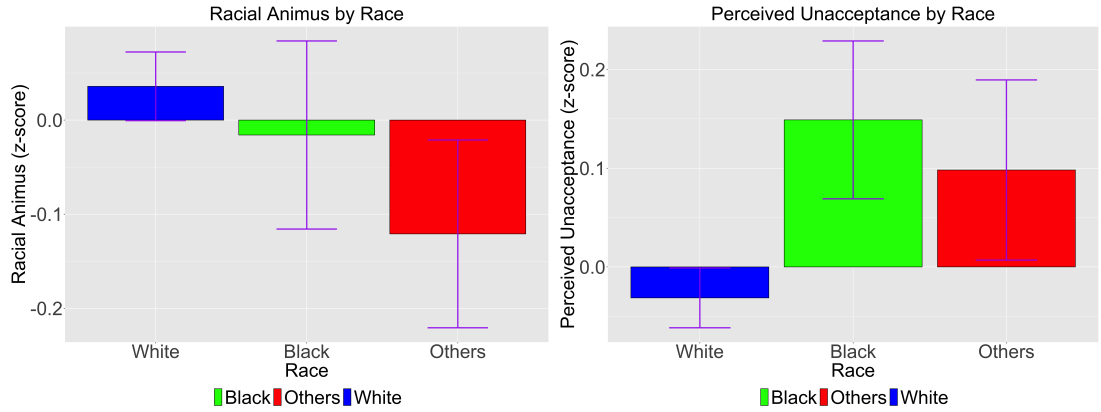
Electronic copy available at: https://ssrn.com/abstract=3958536

Table 2: Sample Balance

Variable	Main Survey	US Population	Variable	Main Survey	US Population
Male	0.45	0.50	Married	0.50	0.49
18-29 years old	0.19	0.24	$0~$38754	0.32	0.25
30-59 years old	0.58	0.57	$38755~$73978	0.31	0.25
60-70 years old	0.24	0.19	$73979~$129066	0.24	0.25
High School or Below	0.37	0.45	$129067+	0.14	0.25
Some College	0.28	0.25	Northeast	0.22	0.17
College	0.36	0.30	Midwest	0.24	0.21
White	0.80	0.78	South	0.38	0.39
Black/African American	0.11	0.14	West	0.15	0.22
Others	0.09	0.07			
Sample Size	2363				

Note: Asian Americans excluded; Household income data come from ASEC CPS 2019 (Flood et al. (2021)); Other U.S. population data come from ACS 2019 (Ruggles et al. (2021)).

Figure 2: The difference in racial animus and perceived unacceptance by racial groups



Note: Others mean non-white, non-black, and non-asian, for example, Hispanic. Whiskers denote the 95% confidence intervals of the group means.

Table 3: COVID-related experience statistics

Variable	N	Mean	Variable	N	Mean
COVID self	2,354	0.079	Job loss	1,318	0.118
COVID family	2,349	0.262	Work face-to-face	1,318	0.458
COVID relative	2,349	0.289	Telework at home	1,318	0.424
COVID friend	2,349	0.390			

sample got infected with COVID and 26% have a family member who did. Among those who had a job before the pandemic, 12% lost their job and 46% had to continue working face-to-face.

Electronic copy available at: https://ssrn.com/abstract=3958536

Table 4: Reduced-Form Evidence for Theory

	Dependent variable:			
	Xenophobic Donation	Xenophobic Petition	(DG) 1(White>Chinese)	(DG) (White-Chinese)
Racial Animus (z-score)	0.122***	0.125***	0.094***	2.435***
	(0.011)	(0.008)	(0.008)	(0.296)
Perceived Unacceptance of Racial Animus (z-score)	−0.166***	−0.056***	−0.030***	−0.546
	(0.014)	(0.010)	(0.010)	(0.355)
Weighted Average of Dependent Variable	0.232	0.098	0.097	-0.138
Observations	2,148	2,148	2,148	2,148
R^2	0.184	0.168	0.087	0.046

Note: DG stands for dictator game. The dependent variable in the third column is whether a respondent shared more with a white American than with a Chinese immigrant. The dependent variable in the fourth column is the difference between the share with a White American and the share with a Chinese immigrant. We use the weight variable provided by a survey firm to match the US representative population.
***p<0.01

5.2 Reduced-Form Evidence

The data confirmed consistent patterns with our theory of xenophobia. Table 4 shows regression coefficients when regressing Sinophobic behaviors on the average z-score index of racial animus and perceived unacceptance of racial animus. Our theory predicts that the Sinophobic behavior would be positively correlated with racial animus and negatively correlated with perceived unacceptance of racial animus. The results confirm this pattern.

We find that the information RCT lowered the perceived unacceptance but the treatment effects on either racial animus or xenophobic actions were statistically insignificant (Table 5, 6).[31] To understand the lack of statistical significance on xenophobic behaviors, we did Monte Carlo simulations explained in Online Appendix G. The evidence suggests that this is likely

[31]This is consistent with our prior that the information will affect the image concern, as described in our AEA RCT registry.

Electronic copy available at: https://ssrn.com/abstract=3958536

Table 5: Information RCT ITT/LATE on v and μ

	Dependent variable:			
	Racial Animus (z-score)	Perceived Unacceptance (z-score)	Racial Animus (z-score)	Perceived Unacceptance (z-score)
	ITT		LATE	
Whether Assigned Treatment	−0.035	−0.061**		
	(0.033)	(0.029)		
Treatment			−0.040	−0.069**
			(0.038)	(0.033)
Observations	2,345	2,154	2,345	2,154

Note : The compliance rate for treatment was 87%. **p<0.05; We use the weight variable provided by a survey firm to match the US representative population.

Table 6: Information RCT ITT and LATE on a

	Dependent variable:			
	Xenophobic Donation	Xenophobic Petition	(DG) 1(White>Chinese)	(DG) (White-Chinese)
	ITT			
Whether Assigned to Treatment	−0.012	−0.003	0.003	0.118
	(0.017)	(0.012)	(0.012)	(0.407)
	LATE			
Treatment	−0.014	−0.004	0.003	0.135
	(0.020)	(0.014)	(0.014)	(0.466)
Weighted Average of Dependent Variable	0.232	0.098	0.097	-0.138
Observations	2,363	2,363	2,363	2,363

Note: The compliance rate for treatment was 87%. We use the weight variable provided by a survey firm to match the US representative population.

to happen if the effect on perceived unacceptance is not big enough: the probability of finding a significant treatment effect on xenophobic actions conditional on finding a significant treatment effect on perceived unacceptance z-score is not that high – between 8% and 21% only – when we use the data generating process that closely matches the estimates (Online

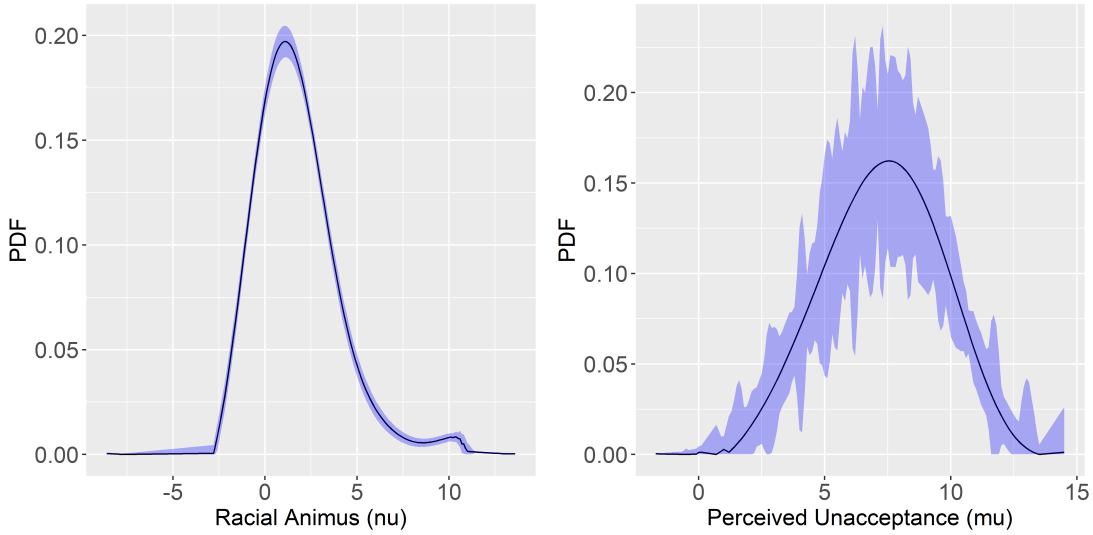
Electronic copy available at: https://ssrn.com/abstract=3958536

Appendix Figure G.1). If the effect on perceived unacceptance were stronger, say three times the current estimates, then the probabilities would go up to a range between 10% and 76%. The probability is biggest for a xenophobic donation, 76%, whose κ parameter estimate is the largest, and the probability is smallest for a dictator game outcome, 10%, whose κ parameter estimate is the smallest.

6 Estimation Results

6.1 Density Estimation

Figure 3: Estimated Density of Racial Animus and Perceived Unacceptance



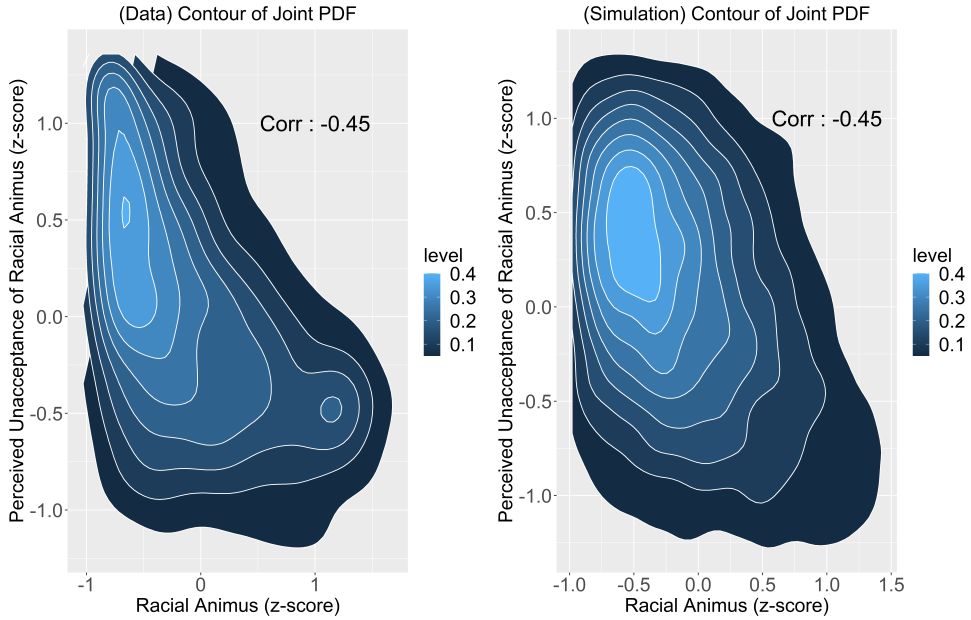
Note: This figure shows the estimated densities of racial animus ν and perceived unacceptance μ using the Li and Vuong (1998) deconvolution kernel estimator. The 95% confidence interval is computed from bootstrapping 100 times and is denoted as a shaded area.

Figure 3 shows the estimated densities of racial animus and perceived unacceptance. The density of racial animus is estimated to be tighter than that of perceived unacceptance. This is likely due to a smaller correlation between proxy measures for perceived unacceptance (Table A.2 in Online Appendix). Another notable feature is that the density of racial animus is skewed to the right, while the density of the perceived unacceptance is symmetric. This means there is a small number of extreme haters, and most people show a mild degree of racial animus. On the other hand, the perception of unacceptance of racial animus is more symmetrically dispersed and is inverted-U shaped.

For brevity, I present the measurement equation parameter estimates and the percentage of signal and noise for each proxy in Table A.3 in the Online Appendix. The signal ratio

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure 4: Model Fit for Joint Density of Racial Animus and Perceived Unacceptance



Note: This figure shows the model fit of the joint density of racial animus v and perceived unacceptance μ.

indicates the informativeness of each proxy about the latent variable and the noise ratio is defined as one minus the signal ratio. Mathematically, the percentage of signal is defined as $\frac{(\alpha^a_{k1})^2 Var(a)}{((\alpha^a_{k1})^2 Var(a)+Var(\epsilon^a_k))}$, and the percentage of noise is defined as $\frac{Var(\epsilon^a_k)}{((\alpha^a_{k1})^2 Var(a)+Var(\epsilon^a_k))}$ for $a \in \{v, \mu\}$. We find the informativeness of each proxy variable varies widely, which is shown by a vast range of signal and noise ratios, ranging from 0.14 to 0.82.

Overall, we have a good model fit. Figure 4 shows the model fit of the joint density of racial animus and perceived unacceptance. Given the Joe copula parameter estimate, the simulated data fits well with the empirical joint density of average z-scores of racial animus and perceived unacceptance. Perceived unacceptance and racial animus are negatively correlated, with a correlation coefficient of -0.45, but are never perfectly correlated. In particular, among people with small racial animus, there is a large dispersion in the perception of unacceptance of racial animus. This means perceived unacceptance and racial animus are distinct constructs. Another notable feature is that people with high racial animus tend to perceive that racial animus is acceptable with a small variance. This may be due to a psychological tendency to have a positive self-image[32]. Finally, Figure A.1 in the Online Appendix shows the model fit of raw proxy variables of racial animus and perceived unacceptance. We have a reasonably good fit.

[32]In the long run, there may be feedback between the evolution of racial animus and perceived unacceptance. However, given our cross-sectional data, studying the dynamic evolution of racial animus and perceived unacceptance is beyond the scope of this study.

Electronic copy available at: https://ssrn.com/abstract=3958536

Table 7: Structural Parameter Estimates

	Parameter Meaning	Xenophobic action Xenophobic Donation	Xenophobic Petition	(DG) 1(White>Chinese)
κ	scale parameter for μ	1.85	0.53	0.25
		(0.25)	(0.07)	(0.06)
c	location parameter for μ	-5.90	0.57	3.48
		(1.56)	(0.26)	(0.21)
β	Gumbel shock scale	8.69	4.00	3.96
		(0.82)	(0.36)	(0.12)
θ	Joe copula parameter		2.08	
			(0.11)	

Note: The standard errors are in parentheses. They are computed by bootstrapping the entire estimation procedure 100 times.

6.2 Structural Estimates

Table 7 shows a subset of structural parameter estimates. The structural parameters (κ, c, β) are allowed to differ by xenophobic actions. All standard errors are reasonably small. Table 8 shows that the model fit for 15 targeting moments in the Indirect Inference estimation is good. Every simulated moment is within the 95% confidence intervals of data moments.

The relative importance of image concern is jointly captured by κ and c, that is, the scale

Table 8: Model Fit For Targeting Moments In Indirect Inference

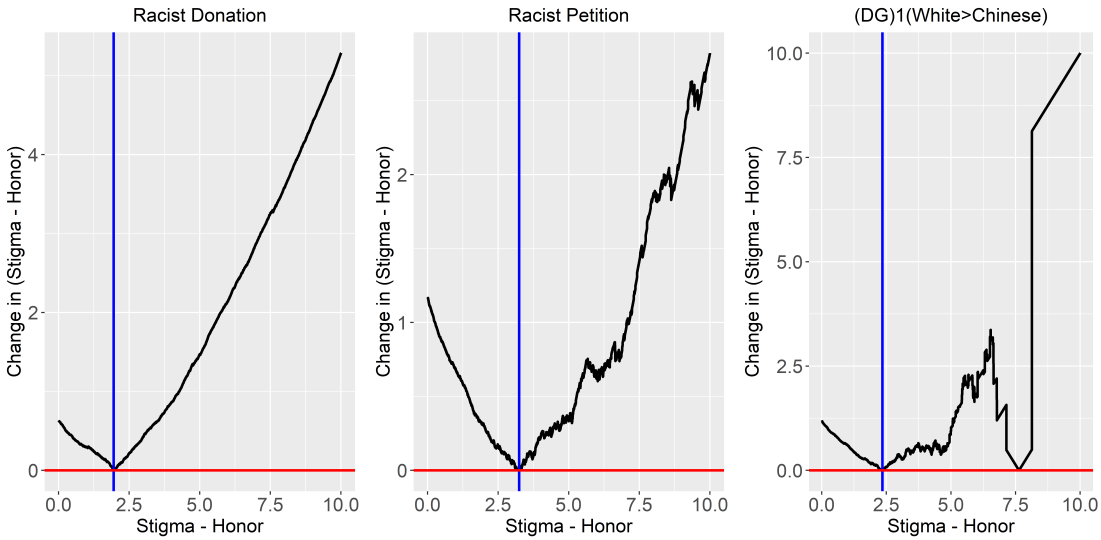
	Xenophobic Donation		Xenophobic Petition		(DG)1(White>Chinese)	
Moments	Data	Model	Data	Model	Data	Model
ξ_0	0.23	0.23	0.09	0.09	0.09	0.09
	[0.21,0.24]		[0.08,0.10]		[0.07,0.10]	
ξ_1	0.11	0.12	0.12	0.12	0.08	0.09
	[0.09,0.13]		[0.09,0.14]		[0.06,0.10]	
ξ_2	-0.17	-0.15	-0.05	-0.06	-0.03	-0.03
	[-0.19,-0.14]		[-0.07,-0.03]		[-0.05,-0.01]	
$P(a=1)$	0.23	0.23	0.09	0.09	0.09	0.09
	[0.21,0.25]		[0.08,0.10]		[0.08,0.10]	
$\widehat{E[v\|a=1]} - \widehat{E[v\|a=0]}$	2.02	1.95	3.41	3.24	2.51	2.35
	[1.79,2.25]		[3.01,3.81]		[2.09,2.93]	

Note : 95% CIs of data moments are in brackets. ξ_0, ξ_1, ξ_2 are regression coefficients in a linear probability model in equation 24.

parameter and a location parameter of μ, and they are important for our counterfactual analysis. They are estimated to vary sizably across different xenophobic actions. This implies the image concern may differ by xenophobic actions, possibly because each xenophobic action

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure 5: Evidence of Multiple Equilibria Under Structural Parameter Estimates



Note: This figure shows that there are multiple equilibria under the structural parameter estimates (for xenophobic behavior during the dictator game, titled '(DG)1(White>Chinese)'). There is no other equilibrium for xenophobic donation or xenophobic petition.

may have different publicity. We have evidence that the image concern is the most important for the xenophobic petition and the least important for the dictator game. This is shown in our counterfactual analysis in Section 7.1 and the Twitter sample analysis in Online Appendix Section F.

In Figure 5, we examine whether there are other equilibria under our estimated structural parameters and find there are. Each panel in Figure 5 shows the change in reputational gain $E[v|a = 1] - E[v|a = 0]$ after applying the fixed point mapping implied by the theory given the current reputational gain. When there is no change, shown as a tangent point on a zero-horizontal line, it means there is an equilibrium corresponding to the reputational gain. In the third panel of xenophobic behavior during the dictator games, we see another equilibrium with much higher reputational gain. For the two other xenophobic actions we consider, we do not find evidence of multiple equilibria.

6.3 Validation

We validate our structural parameter estimates by comparing our model prediction for the RCT treatment effect on xenophobic actions with the reduced-form causal estimates, which were presented in Table 6. To predict the RCT treatment effect using model estimates, we first estimate the densities of racial animus and perceived unacceptance by treatment status, and next, we predict the xenophobic actions for the treated and control group, given the densities

Electronic copy available at: https://ssrn.com/abstract=3958536

by treatment status and the structural parameter estimates. In this computation, we take the reputational gains fixed at the estimated level because we do not expect the participants to take into account the RCT effect on the reputational gain – that is, we compare the short-run prediction estimates with the reduced-form causal estimates. We compute the standard errors of model prediction by bootstrapping 100 times.

Table 9 shows that model predictions for the ITT are close to the reduced-form intention-to-treat (ITT) estimates. We could not reject that these estimates are statistically different. It is reassuring that we can replicate the reduced-form intention-to-treat estimates using our structural parameter estimates.

Table 9: Validation of Structural Parameter Estimates

	ITT (Model)	ITT (Data)	Difference p-value
Xenophobic Donation	0.00	-0.01	0.73
	(0.02)	(0.02)	
Xenophobic Petition	0.00	0.00	0.50
	(0.01)	(0.01)	
Xenophobic Dictator Game	-0.01	0.00	0.26
	(0.00)	(0.01)	

Note: This table compares our model prediction for the intention-to-treat (ITT) estimates for xenophobic behaviors with the reduced-form ITT estimates. The standard errors are inside the brackets. The standard errors for the model prediction are computed by bootstrapping 100 times, and the standard errors for the reduced-form ITTs are from Table 6. The p-values for the differences were computed following Paternoster et al. (1998).

7 Counterfactual Analysis

We make two counterfactual predictions using our estimated structural models. First, we quantify the relative importance of intrinsic motivation versus reputational motivation in reducing xenophobic actions. Second, we predict how COVID infections affect xenophobia both in short and long run.

We set an equilibrium selection rule in our counterfactual analysis due to the presence of multiple equilibria, although we do not need such a rule for estimation. The rule entails choosing the equilibrium with a reputational gain closest to the baseline level. This assumption is reasonable if it is less likely to have an abrupt change in the equilibrium reputational gain.

Assumption 6 (Equilibrium Selection Rule)**.** *We choose an equilibrium whose reputational gain* $E[v|a=1] - E[v|a=0]$ *is closest to the baseline level.*

Electronic copy available at: https://ssrn.com/abstract=3958536

Table 10: Counterfactual Prediction When Shifting Racial Animus and Perceived Unacceptance by the Largest Racial Gap (=0.13 SD) Respectively

		decreasing racial animus (ν)		increasing perceived unacceptance (μ)	
		Holding (stigma - honor) fixed as baseline			
	baseline	p.p. ch	% ch	p.p. ch	% ch
Xenophobic Donation	0.23	-0.46	-2.01	-1.62	-7.10
Xenophobic Petition	0.09	-0.50	-5.60	-0.71	-7.97
Xenophobic Dictator Game	0.09	-0.58	-6.78	-0.36	-4.23
		Updating (stigma - honor) in new equilibrium			
	baseline	p.p. ch	% ch	p.p. ch	% ch
Xenophobic Donation	0.23	-0.51	-2.23	-2.54	-11.15
Xenophobic Petition	0.09	-0.74	-8.35	-1.39	-15.56
Xenophobic Dictator Game	0.09	-0.79	-9.14	-0.56	-6.49

Note: This table shows the counterfactual predictions for shifting the racial animus ν and perceived unacceptance μ by the largest racial gap, that is, the difference between an average white person and an average non-white, non-black, non-asian person (Others) (Figure 2). The top panel shows a short-run prediction holding the reputational gain fixed at the baseline level and the bottom panel shows a long-run prediction when updating the reputational gain to a new level.

7.1 Relative Significance of Racial Animus and Perceived Unacceptance

To account for the relative importance between intrinsic motivation and reputational concern, we make counterfactual predictions when shifting racial animus and perceived unacceptance distribution by the largest racial gap, the difference between the most hostile racial group, white respondents, and the most friendly racial group, other race respondents (Figure 2). To give an example of a potential policy that can change each motivation, increasing interaction between immigrant and native students at the school may help reduce racial animus in the long run. Merlino et al. (2019) provides supportive evidence of this policy in a related context: more interracial interactions during childhood can improve racial attitudes toward people of color. And policies sending strong messages against xenophobia can potentially make people update their belief on perceived unacceptance.

Table 10 shows the counterfactual predictions. The top panel shows the short-run prediction, which holds the reputational gain at the baseline level. The bottom panel shows the long-run prediction, which updates the reputational gain to a new fixed point. For changes in the reputational gains in each counterfactual scenario, please see Table A.4 in the Online Appendix. Note that we picked an equilibrium following our equilibrium selection rule (Assumption 6) since there are multiple equilibria for xenophobic behavior during the dictator games.

Electronic copy available at: https://ssrn.com/abstract=3958536

We find that, for xenophobic actions (xenophobic donation and xenophobic petition) with large relative importance for perceived unacceptance (κ), increasing perceived unacceptance μ is more effective at reducing xenophobic actions both in the short and long run. For the xenophobic dictator game, κ is estimated to be the smallest, and c is estimated to be the largest, and, in this case, reducing the racial animus ν is marginally more effective both in the short and long run. A much larger decrease in xenophobic donation and xenophobic petition action occurs in the long run when we increase perceived unacceptance. This is because the reputational gain becomes bigger (Online Appendix Table A.4), and therefore, marginal agents stop engaging in xenophobic actions due to higher reputational consequences.

We do not claim that the conclusion is generalizable to any context. With a different joint distribution of racial animus and perceived unacceptance $F(\nu, \mu)$ and different relative importance of reputational concern captured by the parameter κ and c, the conclusion may change. The takeaway from our analysis is that the joint density of $F(\nu, \mu)$ and the relative importance parameters κ, c matter for the marginal change in xenophobic action. For example, the fact that there is a thin tail of extreme haters while the perceived unacceptance is rather symmetrically distributed around the median makes the perceived unacceptance a more important margin to reduce most xenophobic actions. The publicity of xenophobic action, which can be captured by κ and c, is crucial as well. For a mostly private xenophobic action, reducing racial animus can be more important. If someone wants to know which margin is more important for reducing xenophobic actions, one should examine how racial animus and perceived unacceptance are distributed in society and consider the publicity of those xenophobic actions.

7.2 Effects of COVID Infection

We examine how COVID infection affects xenophobia and conduct a counterfactual analysis of how equilibrium would change if nobody gets infected with COVID. First, we run quantile regressions (equation 7, 8) and predict the distribution of racial animus ν and perceived unacceptance μ under counterfactuals. We make the conditional independence assumption that conditioning on covariates X, the COVID infection event is plausibly random and is independent of potential outcomes of ν and μ.

The credibility of the conditional independence assumption hinges on how well the controlled covariates include most confounding factors that affect both COVID infection and outcome variables ν and μ. Our choice of covariates is guided by previous literature on the potential source of such confounding factors. First, we control for pre-Pandemic political at-

Electronic copy available at: https://ssrn.com/abstract=3958536

titude by the 2016 presidential election vote[33]. Allcott et al. (2020) has shown that there is a substantial difference in social distancing behavior by political partisanship, which results in differential COVID infection rates. And Cao et al. (2022) found substantial differences in anti-Asian sentiment by political partisanship, so not controlling for political attitude would result in an omitted variable bias.

Second, we control for whether the respondent watched a Fox news during the pandemic because several studies (Simonov et al. (2020), Pinna et al. (2021)) found that Fox News viewers didn't observe safety measures such as wearing masks, social distancing, and getting vaccinated, which resulted in much higher rates of COVID infection. And Fox news viewers are more xenophobic, as confirmed in our survey, so not controlling for Fox news viewership may result in a spurious association between COVID infection and higher racial animus. We exclude this possibility by directly controlling for Fox news viewership.

Third, we control the proxies for pre-pandemic attitudes towards asians (number of close asian friends, whether a spouse is asian, and asian shares in all childhood schools, which are primary, secondary, high school, and college). The Contact hypothesis literature has shown that exposure to a different racial group during childhood changes interracial attitudes (Merlino et al. (2019)). By controlling for pre-pandemic attitudes toward asians, we compare respondents with similar pre-pandemic attitudes toward asians but who had different COVID infection histories.

Fourth, we additionally control for characteristics of social networks right before the pandemic that might be correlated with different observances of health safety measures as well as outcome variables ν and μ. They are the difference in the vote share between Hillary Clinton and Donald Trump in the 2016 presidential election in the residing county and the ethnic fractionalization measure[34] of the residing county. Our motivation to control for the political attitude of neighborhood is similar to a reason why we control individual political attitudes. Our choice of controlling the ethnic fractionalization measure is based on Egorov et al. (2021), which has shown that more ethnically diverse neighborhoods more strictly observed social distancing.

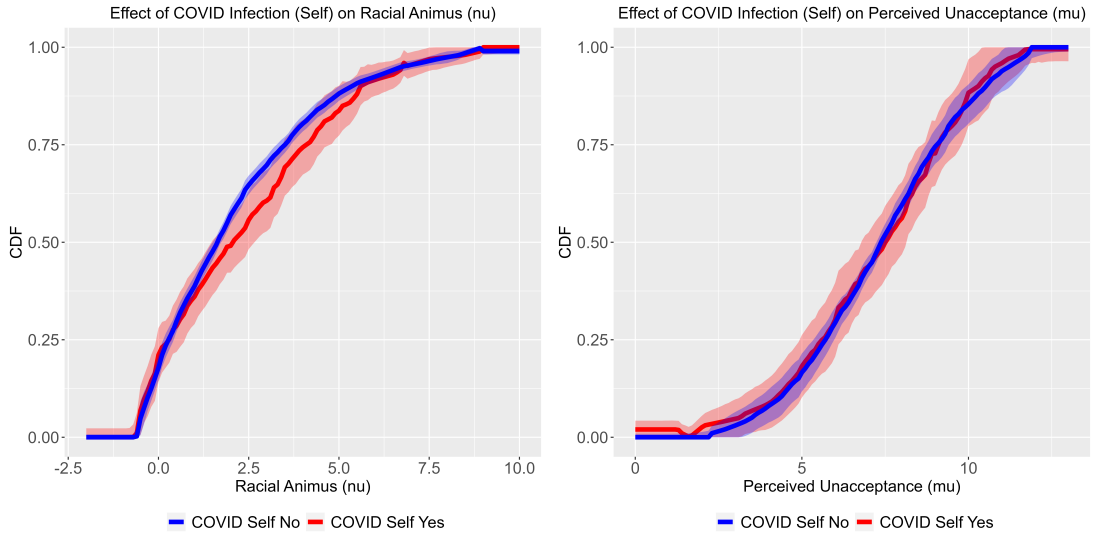
We control for other Pandemic experiences and pre-determined personal characteristics that might confound the effect of COVID. The other Pandemic experiences are whether family members, relatives, or friends have got infected with COVID, whether the respondent lost

[33]Our choice of using the presidential election vote as a proxy for pre-pandemic attitudes toward Chinese immigrants does not require an assumption that all people who voted for Donald Trump are Chinese haters. As long as the distributions of Chinese haters are different by whether someone voted for Donald Trump in 2016, which is supported by a recent empirical study (Cao et al. (2022)), adding the presidential election vote result as an additional control helps explain (partially) the variation in pre-pandemic attitudes toward Chinese immigrants.

[34]This is the probability that randomly chosen two individuals residing in the same county to have different ethnicities.

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure 6: Effect of COVID Infection (Self)



Note: This figure shows the prediction on the CDF of racial animus ν and perceived unacceptance μ if everyone gets infected with COVID (COVID Self Yes) and if everyone does not get infected with COVID (COVID Self No). The shades show a 95% confidence interval around the estimates, which is computed from bootstrapping 100 times. The COVID infection polarizes racial animus, as shown by more mass at both tails for COVID Self Yes. On the other hand, COVID infection does not change the distribution of perceived unacceptance.

a job during the Pandemic, and whether the respondent worked face-to-face during the Pandemic. For pre-determined personal characteristics, we control for race, education, marital status, gender, age, the year 2019 household income quantile groups, employment status in the year 2019, and state fixed effect. Conditional on the above covariates, there is little scope for other unmeasured factors to affect both COVID infection and outcome variables ν and μ.

We find COVID infection polarizes racial animus that is shown by more mass at both tails of the distribution but only minimally changes perceived unacceptance (Figure 6). The shades in Figure 6 indicate the 95% confidence interval, and the CDF of racial animus when no one gets infected with COVID is marginally different from that of when everyone gets infected with COVID. Other COVID-related experiences – whether someone close (family/relative/friend) to a respondent got infected with COVID and changes in work mode during the pandemic (job loss/work face-to-face/telework) – change little racial animus or perceived unacceptance, so we focus on the effect of COVID infection in counterfactual analysis.

Table 11 shows counterfactual predictions. There is no evidence of multiple equilibria in the counterfactual, so this is a unique prediction. In the short run, COVID infection increases xenophobic actions, as shown in the top panel. However, in the long run, the effect of COVID infection is much milder than in the short run, and one action out of three, a xenophobic donation, even decreases in the long run. This is because there is much higher reputational gain from not engaging in these actions (see Table A.5 in the Online Appendix for reputational change in counterfactual scenarios): COVID infection increases the share of extreme haters,

Electronic copy available at: https://ssrn.com/abstract=3958536

Table 11: Counterfactual Predictions for Different COVID Infection (Self) Scenarios

	Holding the (stigma - honor) fixed as baseline			
	COVID (Self) Infection		Yes - No	
	No	Yes	p.p. ch	% ch
Xenophobic Donation	0.23	0.24	1.10	4.84
Xenophobic Petition	0.09	0.10	1.03	12.06
Xenophobic Dictator Game	0.08	0.08	0.60	7.71
	Updating the (stigma - honor) in new eqm			
	No	Yes	p.p. ch	% ch
Xenophobic Donation	0.24	0.22	-1.38	-5.83
Xenophobic Petition	0.13	0.14	0.31	2.36
Xenophobic Dictator Game	0.11	0.11	0.06	0.52

Note: This table shows the counterfactual predictions for when everyone gets infected with COVID and when no one gets infected with COVID. The top panel shows a short-run prediction holding the reputational gain fixed at the baseline level and the bottom panel shows a long-run prediction when updating the reputational gain to a new level.

and therefore, xenophobic actions signal much higher racial animus in the long run. As a result, marginal agents choose to avoid xenophobic behaviors to be distinguished from the extreme haters who engage in xenophobic actions.

8 Conclusion

We present a structural model of xenophobia and estimate our model using newly developed survey instruments to identify motivations behind xenophobic actions. Our survey instruments are essential for identification and estimation because our model can potentially have multiple equilibria. We validate our structural estimation result using the information RCT implemented during the survey.

Our model can be used to quantify the relative importance of racial animus and perceived unacceptance in reducing xenophobic actions. We find that raising perceived unacceptance is more effective than suppressing racial animus at reducing most xenophobic behaviors measured in our survey. The only exception is the dictator game, for which the relative importance of reputational motivation is estimated to be the smallest; this might be because the action in the dictator game is considered rather private.

The reasons why raising perceived unacceptance is more effective for changing xenophobic behaviors in our data are twofold. First, there are more switchers when we raise perceived

Electronic copy available at: https://ssrn.com/abstract=3958536

unacceptance than when we reduce racial animus. This is the short-run effect, and its size depends on the distributional shape of racial animus, perceived unacceptance, and the current location of the indifference line. Second, there is a more significant change in reputational gain from xenophobic (in)action when we shift perceived unacceptance. Higher reputational gain from xenophobic (in)action makes people with moderate racial animus halt xenophobic actions. This is the long-run effect, which determines the long-run outcome.

We get an optimistic prediction for the effect of COVID infection on xenophobia. Although COVID infection increases xenophobic actions in the short run, the increase can be mild in the long run due to the reputational gain from xenophobic (in)action being higher. COVID infections polarize racial animus and increase the number of people with very high racial animus who newly engage in xenophobic behaviors. However, that increased stigma for xenophobic behaviors causes people with moderate racial animus to avoid xenophobic behaviors in the long run over reputational concerns. It would have been difficult to make this long-run prediction if we had no structural model of xenophobia.

Last, our study calls for future work on the determinants of racial animus and perceived unacceptance. Due to the cross-sectional nature of our data, it is infeasible to study how these two motivations form in the first place, nor can we address the possibility of any dynamic feedback effects between racial animus and perceived unacceptance. If there is any psychological bias regarding positive self-image, the racial animus is likely to affect one's perceived unacceptance. Also, if there is homophily in social networks by racial animus, perceived unacceptance may negatively correlate with racial animus, as we observe in our data. Racial animus and perceived unacceptance are important drivers of xenophobic behaviors, so future work on how these two motivations are determined in a dynamic framework will enhance our understanding of how to deter xenophobia.

References

Allcott, H., L. Boxell, J. Conway, M. Gentzkow, M. Thaler, and D. Yang (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of public economics 191*, 104254.

Altonji, J. G. and L. M. Segal (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics 14*(3), 353–366.

Bayer, C., R. Lütticke, L. Pham-Dao, and V. Tjaden (2019). Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk. *Econometrica 87*(1), 255–290.

Electronic copy available at: https://ssrn.com/abstract=3958536

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American economic review 96*(5), 1652–1678.

Berinsky, A. J., M. F. Margolis, and M. W. Sances (2014). Separating the shirkers from the workers? making sure respondents pay attention on self-administered surveys. *American Journal of Political Science 58*(3), 739–753.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. *Handbook of economic field experiments 1*, 309–393.

Bruins, M., J. A. Duffy, M. P. Keane, and A. A. Smith Jr (2018). Generalized indirect inference for discrete choice models. *Journal of econometrics 205*(1), 177–203.

Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review 110*(11), 3522–48.

Butera, L., R. Metcalfe, W. Morrison, and D. Taubinsky (2022). Measuring the welfare effects of shame and pride. *American Economic Review 112*(1), 122–68.

Cao, A., J. M. Lindo, and J. Zhong (2022). Can social media rhetoric incite hate incidents? evidence from trump's "chinese virus" tweets. *NBER Working Paper No. 30588*.

Cibelli, K. (2017). *The Effects of Respondent Commitment and Feedback on Response Quality in Online Surveys*. Ph. D. thesis.

Clifford, S. and J. Jerit (2015). Do attempts to improve respondent attention increase social desirability bias? *Public Opinion Quarterly 79*(3), 790–802.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Delaigle, A. and I. Gijbels (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Computational statistics & data analysis 45*(2), 249–267.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2016). Voting to tell others. *The Review of Economic Studies 84*(1), 143–181.

Dubé, J.-P., X. Luo, and Z. Fang (2017). Self-signaling and prosocial behavior: A cause marketing experiment. *Marketing Science 36*(2), 161–186.

Egorov, G., R. Enikolopov, A. Makarin, and M. Petrova (2021). Divided we stay home: Social distancing and ethnic diversity. *Journal of Public Economics 194*, 104328.

Electronic copy available at: https://ssrn.com/abstract=3958536

Elías, J. J., N. Lacetera, and M. Macis (2019). Paying for kidneys? a randomized survey and choice experiment. *American Economic Review 109*(8), 2855–88.

Flood, S., M. King, R. Rodgers, S. Ruggles, J. R. Warren, and M. Westberry (2021). Integrated public use microdata series, current population survey: Version 9.0.

Glynn, A. N. (2013). What can we learn with statistical truth serum? design and analysis of the list experiment. *Public Opinion Quarterly 77*(S1), 159–172.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of applied econometrics 8*(S1), S85–S118.

Grigorieff, A., C. Roth, and D. Ubfal (2018). Does information change attitudes towards immigrants? representative evidence from survey experiments. *Representative Evidence from Survey Experiments (March 10, 2018)*.

Haaland, I., C. Roth, and J. Wohlfart (2022). Designing information provision experiments. *Journal of Economic Literature (forthcoming)*.

Hu, Y. and G. Ridder (2012). Estimation of nonlinear models with mismeasured regressors using marginal information. *Journal of Applied Econometrics 27*(3), 347–385.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Hubbard, M. L., R. A. Casper, J. T. Lessler, et al. (1989). Respondent reactions to item count lists and randomized response. *Proceedings of the Survey Research Section', American Statistical Association, Washington, DC*, 544–448.

Karing, A. (2019). *Social Signaling and Health Behavior in Low-Income Countries*. Ph. D. thesis, UC Berkeley.

Karlan, D. S. and J. Zinman (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics 98*(1), 71–75.

Kato, K., Y. Sasaki, and T. Ura (2021). Robust inference in deconvolution. *Quantitative Economics 12*(1), 109–142.

Kotlarski, I. (1966). On some characterization of probability distributions in hilbert spaces. *Annali di Matematica Pura ed Applicata 74*(1), 129–134.

Li, T. and Q. Vuong (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis 65*(2), 139–165.

Electronic copy available at: https://ssrn.com/abstract=3958536

Lu, R. and S. Y. Sheng (2022). How racial animus forms and spreads: Evidence from the coronavirus pandemic. *Journal of Economic Behavior & Organization 200*, 82–98.

Merlino, L. P., M. F. Steinhardt, and L. Wren-Lewis (2019). More than just friends? school peers and adult interracial relationships. *Journal of Labor Economics 37*(3), 663–713.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The computer journal 7*(4), 308–313.

Oppenheimer, D. M., T. Meyvis, and N. Davidenko (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology 45*(4), 867–872.

Paluck, E. L., R. Porat, C. S. Clark, and D. P. Green (2021). Prejudice reduction: Progress and challenges. *Annual review of psychology 72*, 533–560.

Paternoster, R., R. Brame, P. Mazerolle, and A. Piquero (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology 36*(4), 859–866.

Pinna, M., L. Picard, and C. Goessmann (2021). Cable news and covid-19 vaccine compliance. *Available at SSRN 3890340*.

Ruggles, S., S. Flood, S. Foster, R. Goeken, J. Pacas, M. Schouweiler, and M. Sobek (2021). Ipums usa: Version 11.0.

Simonov, A., S. K. Sacher, J.-P. H. Dubé, and S. Biswas (2020). The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic. *National Bureau of Economic Research Working Paper*.

Stephan, W. G., O. Ybarra, and G. Bachman (1999). Prejudice toward immigrants 1. *Journal of Applied Social Psychology 29*(11), 2221–2237.

Tessler, H., M. Choi, and G. Kao (2020). The anxiety of being asian american: Hate crimes and negative biases during the covid-19 pandemic. *American Journal of Criminal Justice 45*(4), 636–646.

A Proof

Proof of Proposition 1. The proof can be done in several steps.

Step 1 : Identifying the parameters in measurement equations for proxies

Electronic copy available at: https://ssrn.com/abstract=3958536

We use identifying equations suggested from Cunha et al. (2010). Assumption 2 is necessary to identify these parameters.

$$Var(\nu) = \frac{\sum_{(k,k')} \frac{Cov(Z_1^\nu, Z_k^\nu) Cov(Z_1^\nu, Z_{k'}^\nu)}{Cov(Z_k^\nu, Z_{k'}^\nu)}}{\sum_{(k,k')} 1}, \quad 1 < k, k' < N^\nu, k \neq k' \tag{25}$$

$$Var(\mu) = \frac{\sum_{(g,g')} \frac{Cov(Z_1^\mu, Z_g^\mu) Cov(Z_1^\mu, Z_{g'}^\mu)}{Cov(Z_g^\mu, Z_{g'}^\mu)}}{\sum_{(g,g')} 1}, \quad 1 < g, g' < N^\mu, g \neq g' \tag{26}$$

$$E[\nu] = E[Z_1^\nu] \tag{27}$$

$$E[\mu] = E[Z_1^\mu] \tag{28}$$

$$\alpha_{k1}^\nu = \frac{Cov(Z_1^\nu, Z_k^\nu)}{Var(\nu)} \tag{29}$$

$$\alpha_{g1}^\mu = \frac{Cov(Z_1^\mu, Z_g^\mu)}{Var(\mu)} \tag{30}$$

$$\alpha_{k0}^\nu = E[Z_k^\nu] - \alpha_{k1}^\nu E[\nu] \tag{31}$$

$$\alpha_{g0}^\mu = E[Z_g^\mu] - \alpha_{g1}^\mu E[\mu] \tag{32}$$

$$\sigma_{\epsilon^\nu}^2 = Var(Z_k^\nu) - (\alpha_{k1}^\nu)^2 Var(\nu) \tag{33}$$

$$\sigma_{\epsilon^\mu}^2 = Var(Z_g^\mu) - (\alpha_{g1}^\mu)^2 Var(\mu) \tag{34}$$

Note that $Var(\nu), Var(\mu)$ are overidentified in the model because $Var(\nu) = \frac{Cov(Z_1^\nu, Z_k^\nu) Cov(Z_1^\nu, Z_{k'}^\nu)}{Cov(Z_k^\nu, Z_{k'}^\nu)}$ for $\forall k, k'$ such that $k, k' \neq 1, k \neq k'$ and $Var(\mu) = \frac{Cov(Z_1^\mu, Z_g^\mu) Cov(Z_1^\mu, Z_{g'}^\mu)}{Cov(Z_g^\mu, Z_{g'}^\mu)}$ for $\forall g, g'$ such that $g, g' \neq 1, g \neq g'$. We take an average of expressions obtained from all possible pairs $(k, k'), (g, g')$ to identify $Var(\nu)$ and $Var(\mu)$.

Step 2 : Identifying the joint density $F(\nu, \mu)$

Given the measurement equation parameters identified in the previous step, we can construct the normalized proxies.

$$\widetilde{Z}_k^\nu = \frac{Z_k^\nu - \alpha_{k0}^\nu}{\alpha_{k1}^\nu} = \nu + \widetilde{\epsilon}_k^\nu, \quad k \in \{1, \cdots, N_\nu\}, \quad \widetilde{\epsilon}_k^\nu \overset{i.i.d.}{\sim} N\left(0, \left(\frac{\sigma_{\epsilon^\nu}}{\alpha_{k1}^\nu}\right)^2\right) \tag{35}$$

$$\widetilde{Z}_g^\mu = \frac{Z_g^\mu - \alpha_{g0}^\mu}{\alpha_{g1}^\mu} = \mu + \widetilde{\epsilon}_g^\mu, \quad g \in \{1, \cdots, N_\mu\}, \quad \widetilde{\epsilon}_g^\mu \overset{i.i.d.}{\sim} N\left(0, \left(\frac{\sigma_{\epsilon^\mu}}{\alpha_{g1}^\mu}\right)^2\right) \tag{36}$$

Next, we apply Theorem 1 in Cunha et al. (2010), which is a deconvolution theorem for a vector of random variables – that is, an extension of Kotlarski (1966)'s theorem. To apply Theorem 1 in Cunha et al. (2010), Assumption 3 is necessary; note that Assumption 3 (i) guarantees

Electronic copy available at: https://ssrn.com/abstract=3958536

that the characteristics functions of W_1, W_2 are non-vanishing because measurement errors are assumed to be normal and their characteristics functions are non-vanishing. Applying Theorem 1 in Cunha et al. (2010) guarantees the identification of joint distribution (ν, μ) up to the location. The Assumption 2 (iv) pins down the location of (ν, μ). So the distribution of (ν, μ) is uniquely identified.

Step 3 : Identifying reputational gain $E[\nu|a=1] - E[\nu|a=0]$

Assumption 4 guarantees that both $E[\nu|a=1]$ and $E[\nu|a=0]$ exist. Also, Assumption 1 means there exists a unique reputational gain, $E[\nu|a=1] - E[\nu|a=0]$, that corresponds to the entire data.

$E[\nu|a=1] - E[\nu|a=0]$ is trivially identified from $\{\widetilde{Z}_k^{\nu}, a\}$, a vector of a normalized proxy for ν and an action a.

$$E[\nu|a=1] - E[\nu|a=0] = E[\widetilde{Z}_k^{\nu}|a=1] - E[\widetilde{Z}_k^{\nu}|a=0] \tag{37}$$

Step 4 : Identifying the structural parameters (κ, c, β)

We apply Theorem 1 in Hu and Ridder (2012) for identification. To satisfy Assumption (i) of Theorem 1 in Hu and Ridder (2012), we need to show that our model is identified if the latent variables are observed. To see this, note that the structural parameters (κ, c, β) are identified from the coefficients in the logistic model, $\{\xi_0, \xi_1, \xi_2\}$ and the previously identified term $(E[\nu|a=1] - E[\nu|a=0])$.

$$P(a=1|\nu,\mu) = \frac{exp\left(\frac{\nu-(\kappa\mu+c)(E[\nu|a=1]-E[\nu|a=0])}{\beta}\right)}{exp\left(\frac{\nu-(\kappa\mu+c)(E[\nu|a=1]-E[\nu|a=0])}{\beta}\right)+1} \tag{38}$$

$$\xi_0 = \frac{c(E[\nu|a=1] - E[\nu|a=0])}{\beta} \tag{39}$$

$$\xi_1 = \frac{\kappa}{\beta} \tag{40}$$

$$\xi_2 = \frac{1}{\beta} \tag{41}$$

Assumption 5 assumes $\{\xi_0, \xi_1, \xi_2\}$ are identified when the latent variables and the action $\{\nu, \mu, a\}$ are observed, so this implies Assumption (i) of Theorem 1 in Hu and Ridder (2012) holds. Assumption (ii) of Theorem 1 is trivially satisfied because the measurement error distribution is assumed to be normal in our model. Assumption (iii) is satisfied from the functional form of the logistic model. So Theorem 1 in Hu and Ridder (2012) applies and the structural parameters $\{\kappa, c, \beta\}$ are identified when we observe $\{a, \{Z_k^{\nu}\}_{k=1}^{N_\nu}, \{Z_g^{\mu}\}_{g=1}^{N_\mu}\}$. □

Electronic copy available at: https://ssrn.com/abstract=3958536

B Details on the Survey Design

We carefully designed our survey to ensure high-quality responses.

First, we worded our survey invitation and consent form carefully to avoid selective participation by anti-Chinese racial animus or perceived unacceptance of racial animus. The invitation email did not mention keywords, such as 'anti-Chinese' or 'xenophobia'. Instead, the email invitation started by saying, "New Survey Available!" in the headline, and the email body said "(NAME), you've been pre-qualified to participate in a survey. This survey is only available for a short time, so please respond ASAP!" In the consent form, we described the purpose of our survey vaguely to hide the specific survey topic without deceiving respondents. We said, "The purpose of this survey is to understand the social preferences of people living in the US". We hid our names in the consent form and introduced ourselves as a "non-partisan group of researchers" from University, as our names signal Asian ethnicity and knowing that the research team members are Asians may contaminate the responses.

Second, we asked respondents explicitly at the beginning of the survey to commit to reading the survey carefully and providing honest responses to the best of their ability. Specifically, we said, "You have been selected to represent a portion of the US population. The results from the survey can influence political decisions and thus affect the lives of many people. In order for the information from this research to be the most helpful, it is important that you try to be as accurate, complete, and honest as possible with your answers. To do this, it is important to think carefully about each question, search your memory, and take time in answering. Are you willing to do this?". Cibelli (2017) showed that such an explicit commitment improves the quality of the online survey. We exclude those who refuse to commit to these standards.

Third, we included several quality-check questions to screen out participants paying little attention to our survey and to make participants more attentive throughout the survey. This was recommended by Berinsky et al. (2014), who proved multiple screener questions are effective at improving the quality of online surveys. We inserted two screener questions before important survey blocks which measure key variables (Figure B.1). You can find the two screener questions in Online Appendix C.2. The first screener question pretended to be a question about current feelings, but we asked respondents to check only "None of the above" to prove that they are attentive. We inserted the first screener question just before our information RCT treatment which was followed by questions about racial animus and perceived unacceptance. This was to make participants more attentive during the RCT treatment. The control group did not watch the information RCT treatment video. Instead, they started answering about their racial animus and perceived unacceptance of racial animus right after the first attention check question. The second screener question was masked as a question about an electronic device used to participate in the survey, but we asked respondents to check

Electronic copy available at: https://ssrn.com/abstract=3958536

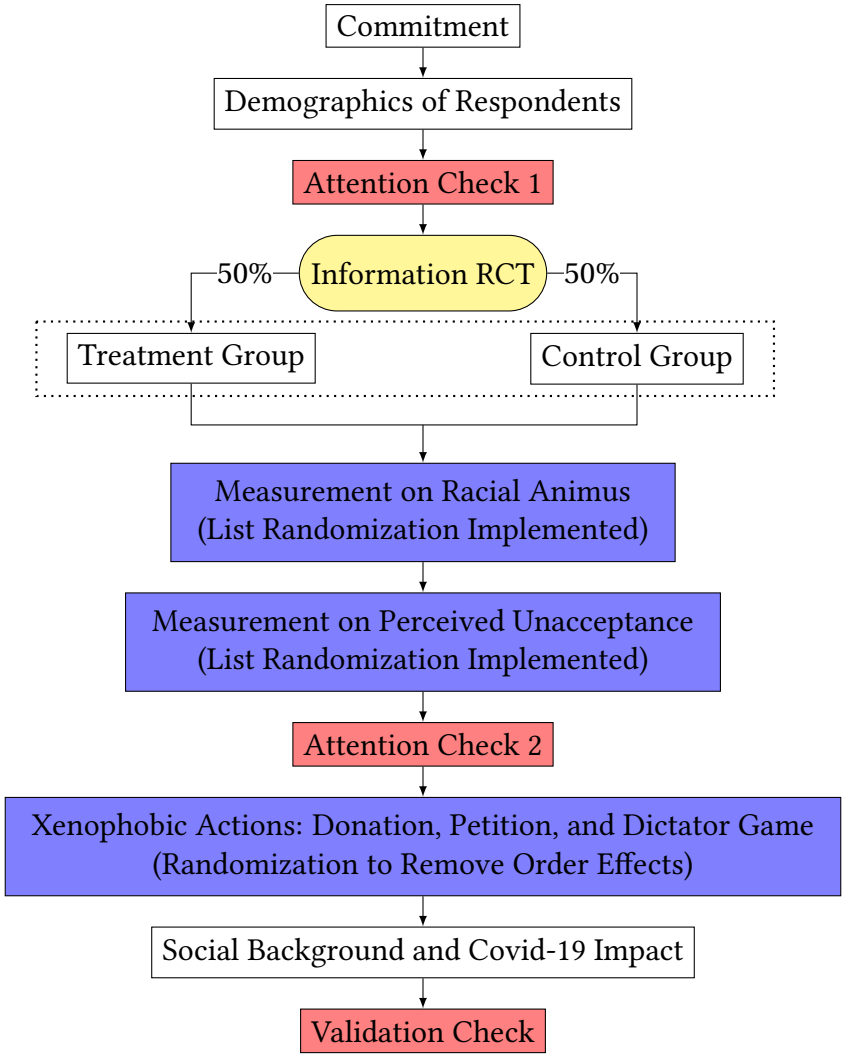


Figure B.1: Survey Flow

"Other." We inserted this question before measuring xenophobic behaviors.

Table B.1 shows the pass rates regarding our attention-check questions. Roughly, 49% to 65% of people passed the attention check questions. Our pass rates are within the range of what people have found elsewhere in the literature (Clifford and Jerit (2015), Oppenheimer et al. (2009), Berinsky et al. (2014)) : Clifford and Jerit (2015) reported only 38% passed their first attention check item and 64% passed the second attention check conditional on passing the first attention check. Oppenheimer et al. (2009) found 54% passed their screener question. Berinsky et al. (2014) showed the pass rates ranged between 59% and 76% for various screener questions.

Fourth, we included a question at the end of the survey asking whether the survey looked biased in favor of or against Chinese immigrants to detect any surveyor demand effect. We dropped a small number of respondents (13%, Table B.1) who answered that the survey looked

Electronic copy available at: https://ssrn.com/abstract=3958536

Table B.1: Dropping low-quality responses

	Try	Pass	Pass rate (%)
Attention Check 1	10641	5187	48.75
Attention Check 2	5187	3372	65.01
Surveyor Demand Check	2723	2363	86.78
Final Sample Size		2363	

biased in either direction or refused to answer this question[35] because their responses may not be honest. For robustness, we repeat our analysis by including the sample who reported the bias in our survey. We confirmed most results remain qualitatively similar.[36]

Dropping potentially low-quality responses brings merit of improving online survey quality but there is a trade-off of selecting a sample based on characteristics that may be relevant to our question. We investigate the degree of potential selection in our sample by adding attention checks and a surveyor demand effect screener and find there are small but significant differences. In Online Appendix E, we show descriptive statistics about the screened sample. The median person who failed to pass either the attention-check question or surveyor demand effect screener spent 1-13 seconds (0.6 - 7% of median) less to respond to racial animus and perceived unacceptance questions and showed 0.1 standard deviations higher racial animus and 0.2 standard deviations lower perceived unacceptance. Therefore, our estimation sample can be less xenophobic than the population and our sample means of xenophobic behaviors can be understood as a lower bound of the true ones.

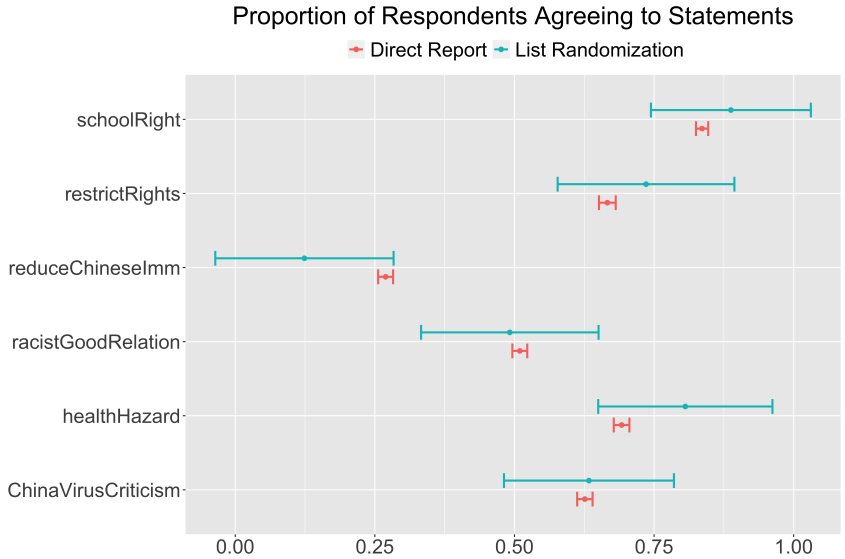
Fifth, we included a battery of List randomization questions to assess social desirability bias. List randomization assigns respondents into either control or treatment groups. The control group was asked to report how many statements out of N neutral statements they agree with, and the treatment group answered a similar question but out of the same N neutral statements plus one extra sensitive statement. The difference between the average response of the control group and the treatment group reveals the fraction of people who agree to the sensitive statement plausibly without social desirability bias. This is because respondents do not have to specify which statement, including a sensitive one, they agree to. If the share of people who agree to the sensitive statement recovered from List randomization is statistically different from the share of people who agree to the sensitive statement in a direct question, it means there is a bias in the direct question, most likely due to social desirability. Each treatment group received one extra sensitive statement about either anti-Chinese animus or perceived unacceptance of racial animus. We carefully chose the neutral questions to avoid the large variance and potential bias in the List randomization responses, which are discussed

[35] 4% answered our survey looked biased against Chinese immigrants, 8% said our survey looked in favor of Chinese immigrants. 1% refused to answer this question, and they were also dropped from the analysis.

[36] This result is available upon request.

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure B.2: Test of Social Desirability Bias using List Randomization



Note: This figure shows the social desirability test for statements about racial animus v and perceived unacceptance μ. We used statements that do not show evidence of social desirability bias.

as a common weakness of the List randomization method (Glynn (2013), Hubbard et al. (1989)). Specifically, we investigated the 2018 ACS data to construct the neutral statements which give the smallest variance in responses and a good mix of prevalent and rare behaviors to prevent floor or ceiling bias. The chosen four neutral statements are "I am a veteran," "I am living with at least one sibling in this household," "I have a smartphone," and "I have health insurance coverage (of any kind, either public or private)." [37]

We used survey instruments for racial animus v or perceived unacceptance μ which do not show evidence of social desirability bias from the List randomization. Figure B.2 compares the shares of people who agreed to a statement about either racial animus v or perceived unacceptance μ from a direct question with the ones from a List randomization question. The figure shows a 95% confidence interval around the point estimates[38]. If the share from a List randomization question is not statistically different from the share from a direct question, there is no evidence of social desirability bias. We found some evidence of social desirability bias from statements included in our survey but we excluded them from our analysis so as not to contaminate the results with social desirability bias as much as possible. See Table A.6 for the statements excluded from our analysis.

Sixth, we randomized the order of choice options in xenophobic behavior measures and the order of the identity of the sequential dictator game partners to remove any order effect.

[37] Our neutral statements are similar to the ones used in Karlan and Zinman (2012).

[38] The confidence intervals from List randomization questions are much wider than the ones from direct questions. However, this is not an error. The high variance is common in List randomization estimates as well known in the literature (Hubbard et al. (1989)).

Electronic copy available at: https://ssrn.com/abstract=3958536

The earlier presented choice option might be implicitly understood as the desirable choice, or some respondents might have a tendency to check options that are presented either earlier or later. By randomizing the choice options, we removed such an order effect on average. Similarly, when we repeated the dictator game sequentially to measure altruism toward partners of different ethnicity, we randomized the order of the partners to mitigate any order effect.

Electronic copy available at: https://ssrn.com/abstract=3958536

FOR ONLINE PUBLICATION: “Structural Analysis of Xenophobia" by Huan Deng and Yujung Hwang

A Additional Tables and Figures

Table A.1: Internal consistency of proxies for ν

	acceptance	warmth	schoolRight	healthHazard	dislike
acceptance					
warmth	0.76***				
schoolRight	0.51***	0.42***			
healthHazard	0.48***	0.40***	0.48***		
dislike	0.68***	0.56***	0.41***	0.35***	
hatred	0.61***	0.52***	0.37***	0.30***	0.75***

Note: This table shows the correlation between proxies for ν. As consistent with a factor model, every proxy is highly correlated with each other. $***$ means the p-value is less than 0.1%.

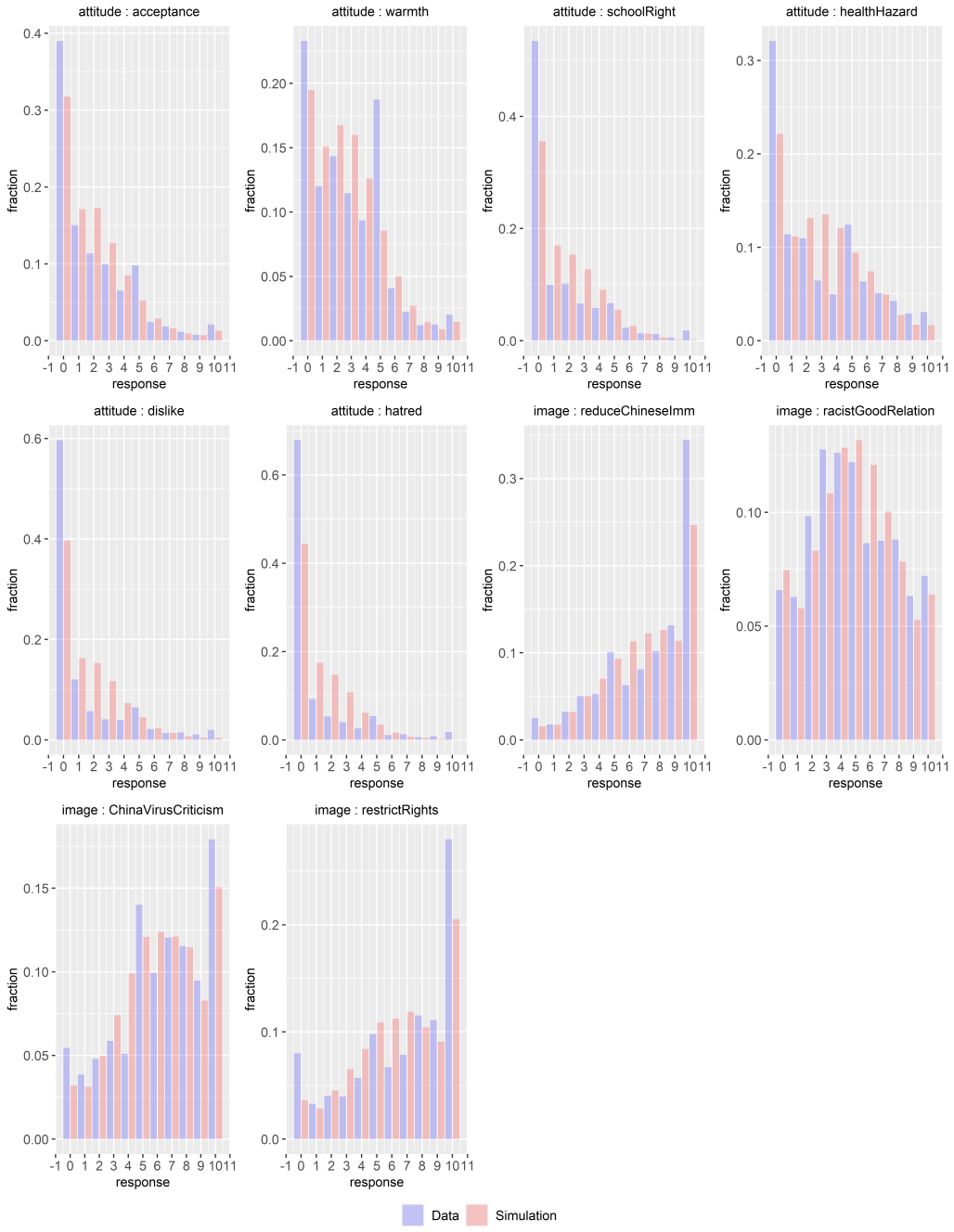
Table A.2: Internal consistency of proxies for μ

	reduceChineseImm	racistGoodRelation	ChinaVirusCriticism
reduceChineseImm			
racistGoodRelation	0.30***		
ChinaVirusCriticism	0.27***	0.19***	
restrictRights	0.24***	0.13***	0.26***

Note: This table shows the correlation between proxies for μ. As consistent with a factor model, every proxy is highly correlated with each other. $***$ means the p-value is less than 0.1%.

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure A.1: Model Fit for Raw Proxies of Racial Animus and Perceived Unacceptance



Note: This figure shows the model fit for raw proxies of racial animus and perceived unacceptance.

Electronic copy available at: https://ssrn.com/abstract=3958536

Table A.3: Measurement Equation Parameter Estimates and % Signal and % Noise

	α_0	α_1	$\sigma^2_{\epsilon^\nu_k}, \sigma^2_{\epsilon^\mu_g}$	% Signal	% Noise
acceptance	0	1	1.03	0.82	0.18
	(0)	(0)	(0.14)	(0.02)	(0.02)
warmth	0.94	0.92	1.82	0.69	0.31
	(0.06)	(0.02)	(0.15)	(0.02)	(0.02)
schoolRight	0.43	0.58	3.93	0.3	0.7
	(0.07)	(0.04)	(0.22)	(0.03)	(0.03)
healthHazard	1.59	0.69	6.71	0.25	0.75
	(0.1)	(0.04)	(0.3)	(0.03)	(0.03)
dislike	-0.24	0.83	2.7	0.55	0.45
	(0.06)	(0.03)	(0.21)	(0.03)	(0.03)
hatred	-0.26	0.68	2.75	0.45	0.55
	(0.05)	(0.03)	(0.21)	(0.03)	(0.03)
reduceChineseImm	0	1	4.68	0.41	0.59
	(0)	(0)	(0.49)	(0.06)	(0.06)
racistGoodRelation	-0.45	0.73	6.41	0.21	0.79
	(0.78)	(0.11)	(0.35)	(0.04)	(0.04)
ChinaVirusCriticism	1.18	0.69	7.19	0.18	0.82
	(0.67)	(0.09)	(0.31)	(0.03)	(0.03)
restrictRights	1.65	0.68	9.04	0.14	0.86
	(0.66)	(0.09)	(0.42)	(0.03)	(0.03)

Note: This table shows the estimates for measurement equation parameters in equation 5, 6, and the percentage of signal and noise from each proxy. The anchor variables are 'acceptance' and 'reduceChineseImm', whose α_0 is normalized to 0 and α_1 is normalized to 1. The standard errors computed from bootstrapping the sample 100 times are in parenthesis. The percentage of signal is defined as $\frac{(\alpha^a_{k1})^2 Var(a)}{((\alpha^a_{k1})^2 Var(a)+Var(\epsilon^a_k))}$, and the percentage of noise is defined as $\frac{Var(\epsilon^a_k)}{((\alpha^a_{k1})^2 Var(a)+Var(\epsilon^a_k))}$ for $a \in \{\nu, \mu\}$.

Table A.4: Reputational Gain, $E^*[\nu|a=1] - E^*[\nu|a=0]$, in Counterfactual 7.1

	baseline	shifts racial animus ν by 0.13 SD counterfactual	shifts perceived unacceptance μ 0.13 SD counterfactual
Xenophobic Donation	1.95	1.96	2.05
Xenophobic Petition	3.24	3.31	3.38
Xenophobic Dictator Game	2.35	2.37	2.37

Note: This table shows the reputational gain in Section 7.1. Reputational gain is higher when perceived unacceptance μ is shifted. This is why shifting perceived unacceptance μ is more effective at reducing most xenophobic behaviors in the long run.

Electronic copy available at: https://ssrn.com/abstract=3958536

Table A.5: Reputational Gain, $E^*[v|a=1] - E^*[v|a=0]$, in Counterfactual 7.2

		COVID (Self) Infection	
		Yes	No
	baseline	counterfactual	counterfactual
Xenophobic Donation	1.95	2.21	1.92
Xenophobic Petition	2.67	2.78	2.66
Xenophobic Dictator Game	2.16	2.24	2.15

Note: This table shows the reputational gain in Section 7.2. COVID infection polarizes racial animus and raises the reputational gain when abstaining from xenophobic behaviors. This is why the effects of COVID infection on xenophobic behaviors, in the long run, are smaller than those in the short run.

The statements not included in the analysis because of the evidence of social desirability bias are summarized in Table A.6

Table A.6: Survey Instruments Not Included In Our Analysis

Anti-Chinese Racial Animus	
Question	Please indicate your agreement with each of the following statements concerning immigration from China. (Slider scale from 0(strongly disagree) to 100(strongly agree))
Statement	• Chinese immigrants get more from this country than they contribute. • Chinese immigrants are contributing to the increase in crime in the US.
Perceived Unacceptance of Racial Animus	
Question	Please indicate your agreement with each of the following statements concerning the attitudes of **typical American citizens (below referred to as "people").** (Slider scale from 0(strongly disagree) to 100(strongly agree))
Statement	• Expressing hatred against Chinese immigrants harms one's social reputation. • People will appreciate it if someone states, "Chinese immigrants bring no good to the US!" • It is socially acceptable to petition to deport Chinese immigrants if they have any Chinese government connections.

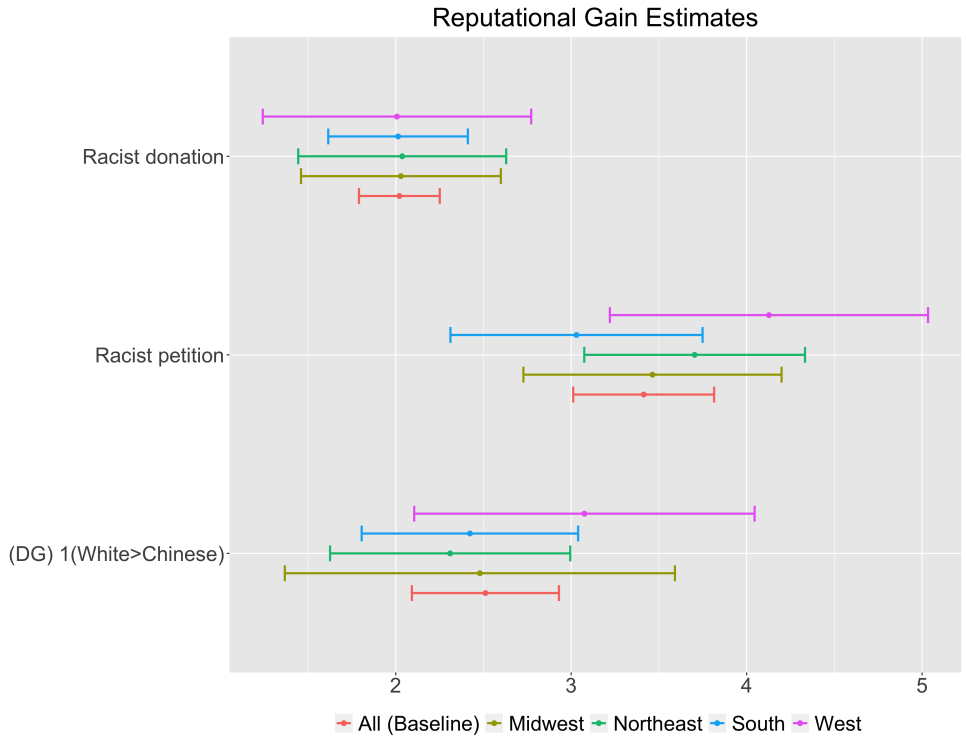
Note: This table shows the list of survey instruments not included in our analysis because of the evidence of social desirability bias.

Electronic copy available at: https://ssrn.com/abstract=3958536

B Robustness Check on a Single Equilibrium Assumption

This section reports supportive evidence of Assumption 1. We examined whether the reputational gains vary across different social networks proxied by US regions and neighborhoods' dominant political attitudes. And we find they are not statistically different across different social networks in our reasonably sized survey sample: Figure B.1 shows that the reputational gains from different US regions are not statistically different from the reputational gain estimated from all regions. Figure B.2 repeats the same exercise but by neighborhoods' dominant political attitudes – whether the majority in the neighborhood voted for either Joe Biden or Donald Trump in the year 2020 presidential election – and finds a similar conclusion. Based on this evidence, we assume that the whole data was generated from a single equilibrium.

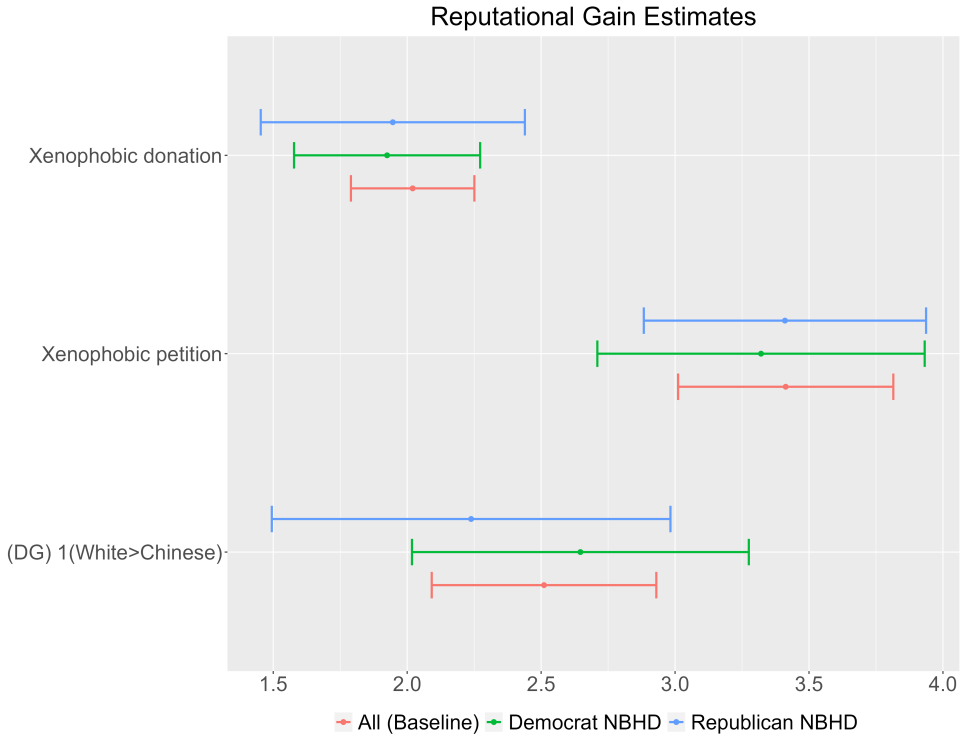
Figure B.1: Robustness checks whether equilibria can be different across the US regions



Note: This figure is to check whether the reputational gains are different across different US regions. We found the reputational gains are not statistically different across the US regions. This supports our Assumption 1 that the entire data corresponds to a single equilibrium.

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure B.2: Robustness checks whether equilibria can be different across the different neighborhood political attitude types



Note: This figure is to check whether the reputational gains are different across neighborhoods of different political attitudes. We coded the neighborhood is "Republican"/"Democrat" if the majority vote in the precinct that the neighborhood belongs to was for "Donald Trump"/"Joe Biden" in 2020 presidential election. The election data was downloaded from https://github.com/TheUpshot/presidential-precinct-map-2020. We found the reputational gains are not statistically different across neighborhoods of different political attitudes. This supports our Assumption 1 that the entire data corresponds to a single equilibrium.

C Survey Questionnaire

This section shows key questions in our survey. For full survey questionnaire screenshots, see our online appendix. You can take our survey from the following link.
https://jhukrieger.co1.qualtrics.com/jfe/form/SV_0wFTvyUFb9nP1NY

C.1 Soft Commitment

We inserted a soft commitment question at the beginning of the survey, following a recommendation by Cibelli (2017).

Electronic copy available at: https://ssrn.com/abstract=3958536

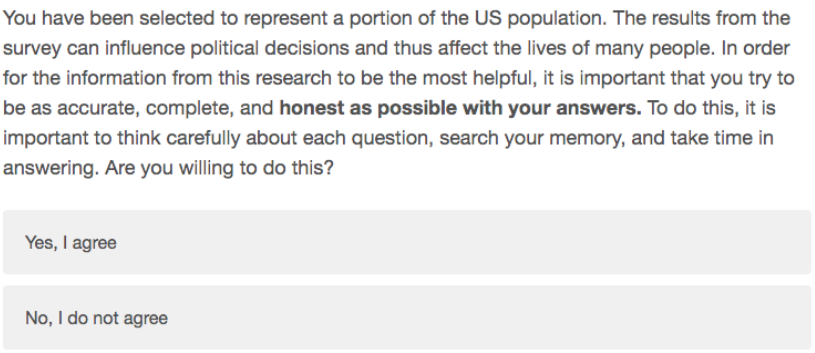


Figure C.1: Soft commitment question

C.2 Attention Check Screener Questions

We included two attention check questions. The first attention check question asks about current feelings, but careful readers will choose the 'None of the above' option only as requested in the question. The second attention check question asks about the device used for the survey, but careful readers will choose 'Other' as requested.

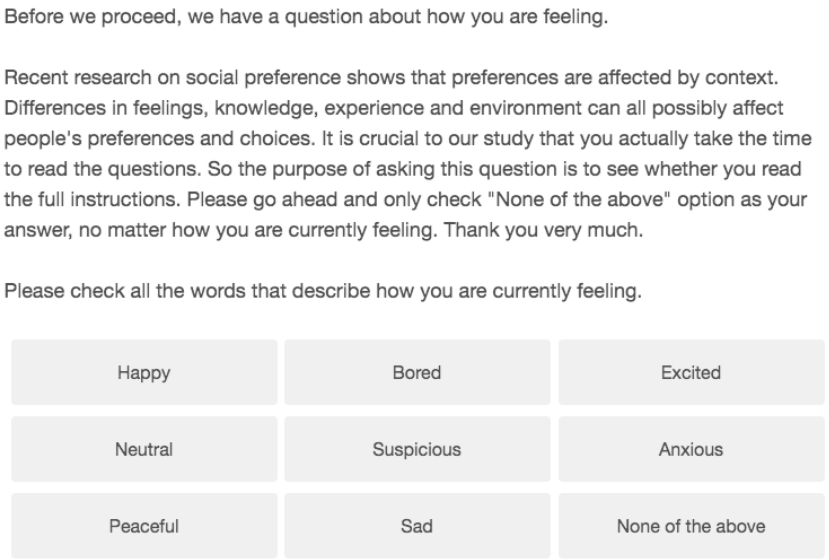


Figure C.2: First attention check question

Electronic copy available at: https://ssrn.com/abstract=3958536

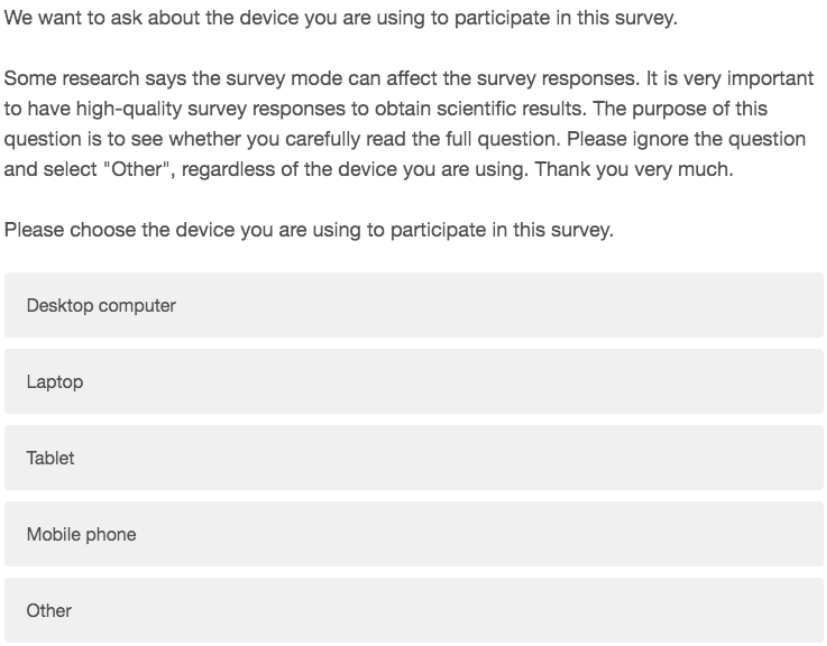


Figure C.3: Second attention check question

C.3 Surveyor Demand Effect Question

We inserted a question at the end of our survey to ask whether participants found our survey biased in favor of or against Chinese immigrants. We dropped the sample who answered our survey looked biased in either direction because their responses may not be honest.

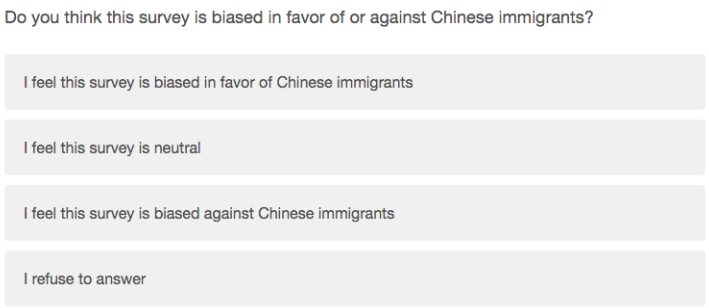


Figure C.4: Question to Check Any Surveyor Demand Effect

C.4 Measurement of Sinophobic Behavior

This section explains our survey questions to measure xenophobic actions.

C.4.1 Donation question

We gave a short description of two organizations with opposing attitudes toward Chinese students and scholars. The donation to Organization A is coded as a Sinophobic behavior.

Electronic copy available at: https://ssrn.com/abstract=3958536

The order of the donation choice was randomized to remove any order effect and surveyor demand effect.

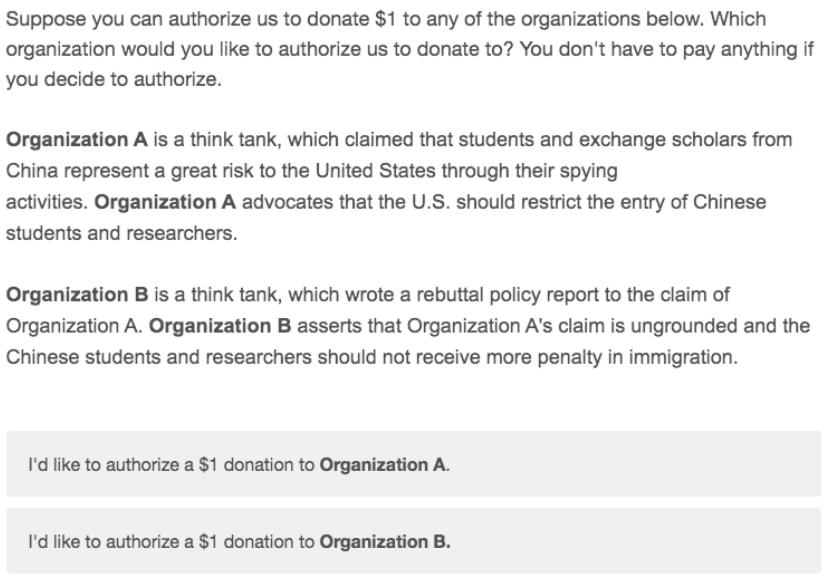


Figure C.5: Donation question to measure Sinophobic behavior

C.4.2 Petition question

We asked if participants wanted to sign any of the petitions below. The decision to sign Petition 2 was coded as a Sinophobic behavior. The order of the petition choice option was randomized to remove any order effect and surveyor demand effect.

Electronic copy available at: https://ssrn.com/abstract=3958536

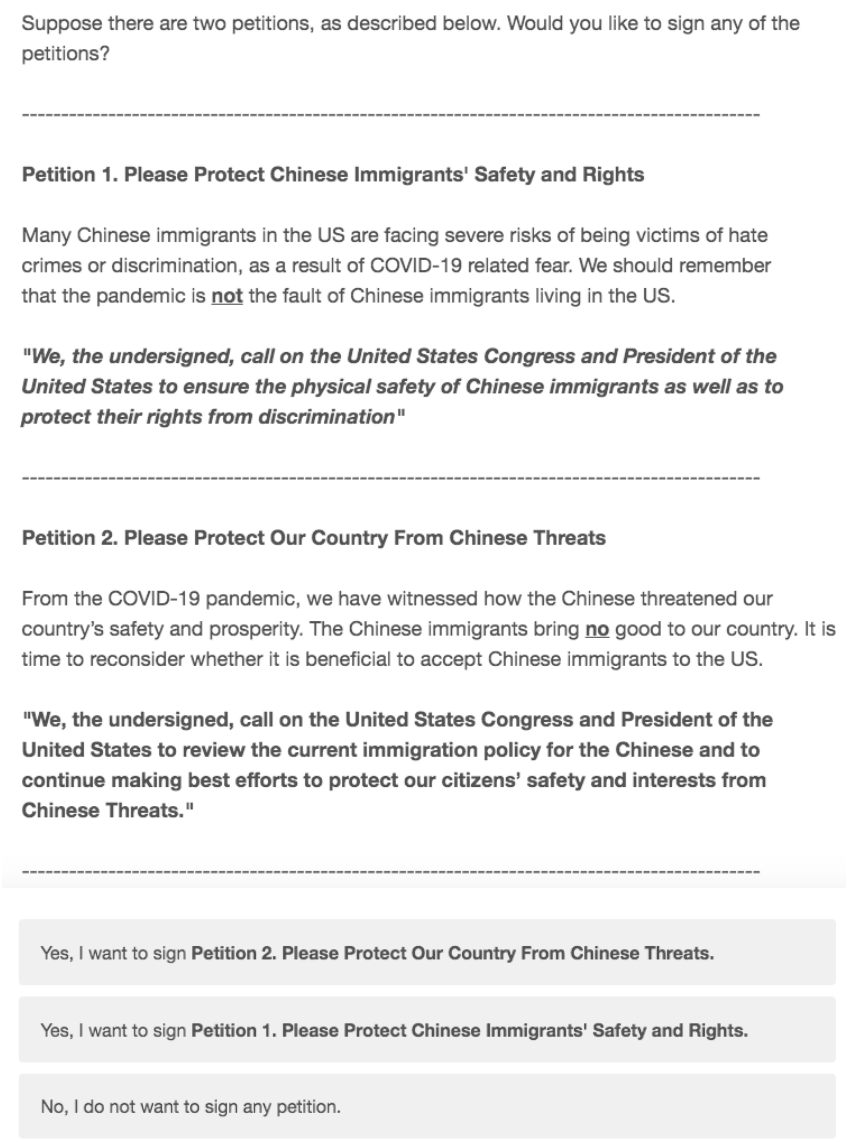


Figure C.6: Petition question to measure Sinophobic behavior

C.4.3 Dictator game

Every participant played a dictator game twice with a Chinese immigrant and a White American. We randomized the order of the partners to remove any order effect.

We coded as a Sinophobic behavior if a participant shared more money with a White American. We use this dummy variable as the main measure of xenophobic behavior instead of the share difference because our model is to explain a discrete action. For robustness, in Table 4, we include the result using the share difference as an outcome variable and obtain qualitatively similar results to when we use a dummy variable as an outcome variable.

Electronic copy available at: https://ssrn.com/abstract=3958536

Now, you will be **randomly** matched with **two people** recruited for this study and you will play a game **twice** with your matched partners. All of your partners are currently living in the US.

You may receive extra rewards based on your responses.

Figure C.7: Introduction to the dictator game

This is your first game. You are matched with the following person.

Name: Haozheng

You are given a lottery to win an extra reward of 100SB (=$1), which can be divided between you and your partner. **10% of survey participants will win the lottery.**

If you win the lottery, how much would you like to give to your partner? If you win, you will be **actually** paid 100SB net of your answer. For example, if you give 50SB to your partner and if you are selected, then you will be paid 50SB (=100SB - 50SB).

Your answer will not affect your probability of winning the lottery.
Please move the slider below to enter your amount to **give** to your partner.

0 10 20 30 40 50 60 70 80 90 100

Amount to your partner (in SB)

Figure C.8: Dictator game with a Chinese immigrant

Electronic copy available at: https://ssrn.com/abstract=3958536

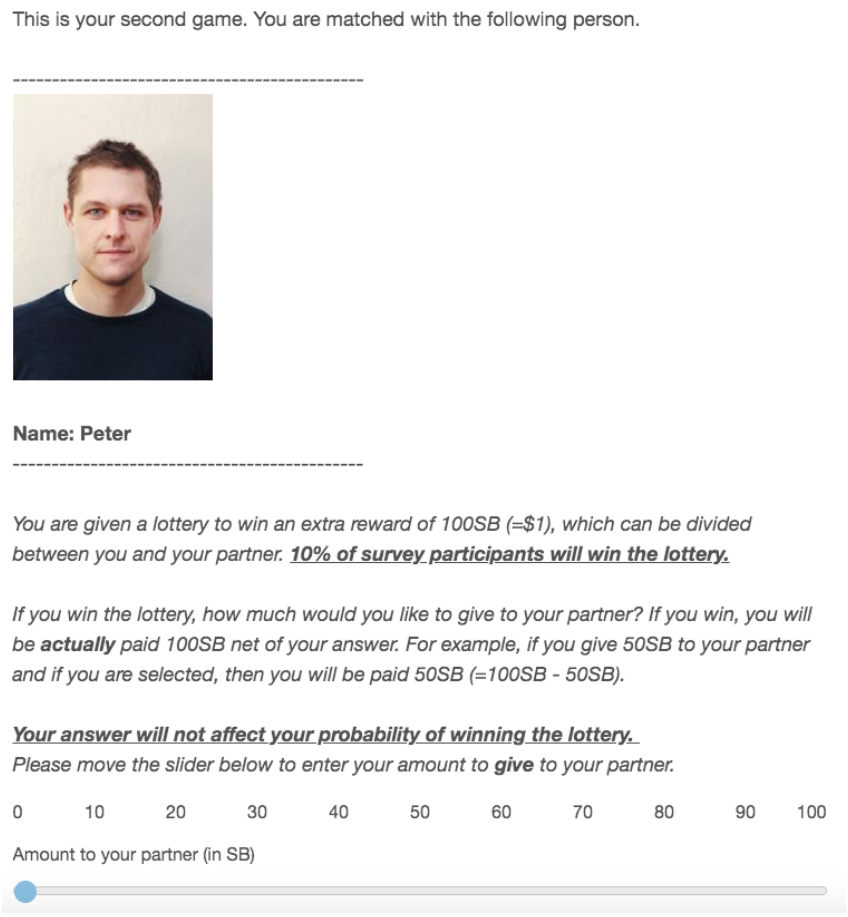


Figure C.9: Dictator game with a White American

C.5 Twitter data

We offered a lottery to win an extra $20. It was impossible to tell the exact winning probability because we can't predict how many people would provide a Twitter username. However, we said that we would randomly select five winners, and we planned to invite 3000 participants to our survey. In the end, we invited less than 3000 participants.

Electronic copy available at: https://ssrn.com/abstract=3958536

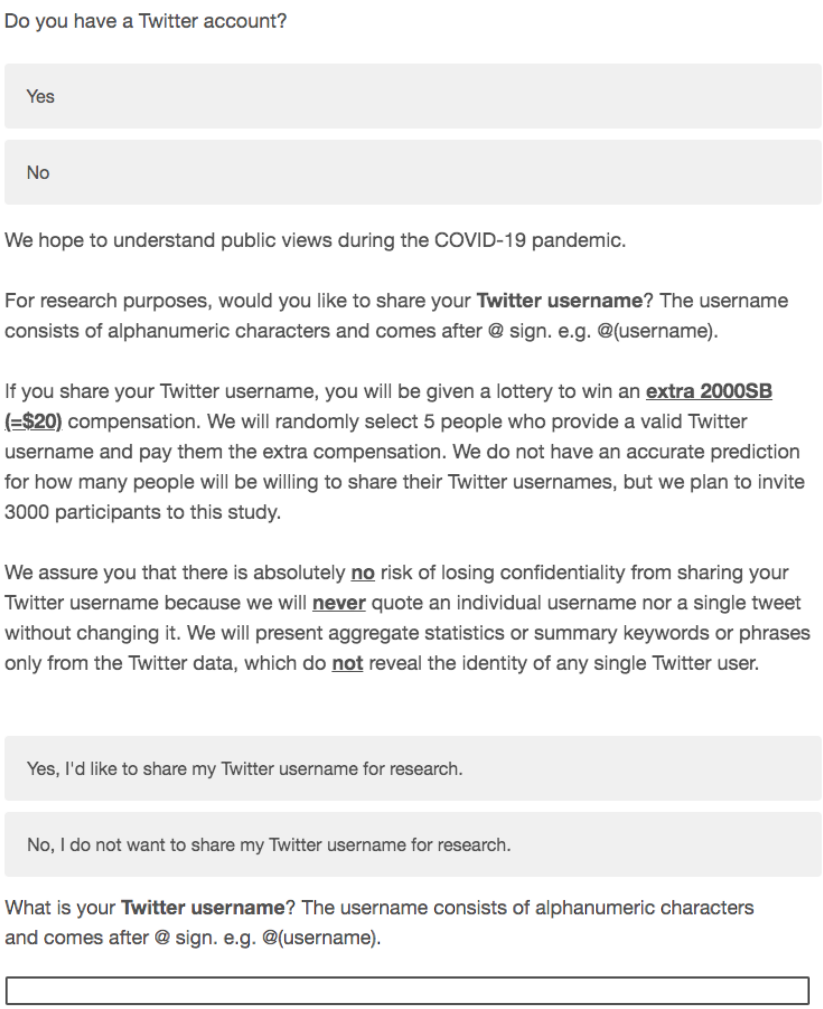


Figure C.10: Questions to collect Twitter usernames

D Robustness Check about RCT Treatment Effect on Social Desirability Bias on Sample Attrition

We examined whether RCT treatment changes social desirability bias later when respondents answer questions about either racial animus or perceived unacceptance of racial animus. We do not find evidence of change in social desirability bias by RCT treatment status. Figure D.1 repeats the test described in Section B by RCT treatment. For both treated and control groups, the means from the List randomization are not statistically different from the means from the direct report. This allows us to interpret the difference in proxy responses stemming from the difference in latent variables instead of the change in measurement errors.

Electronic copy available at: https://ssrn.com/abstract=3958536

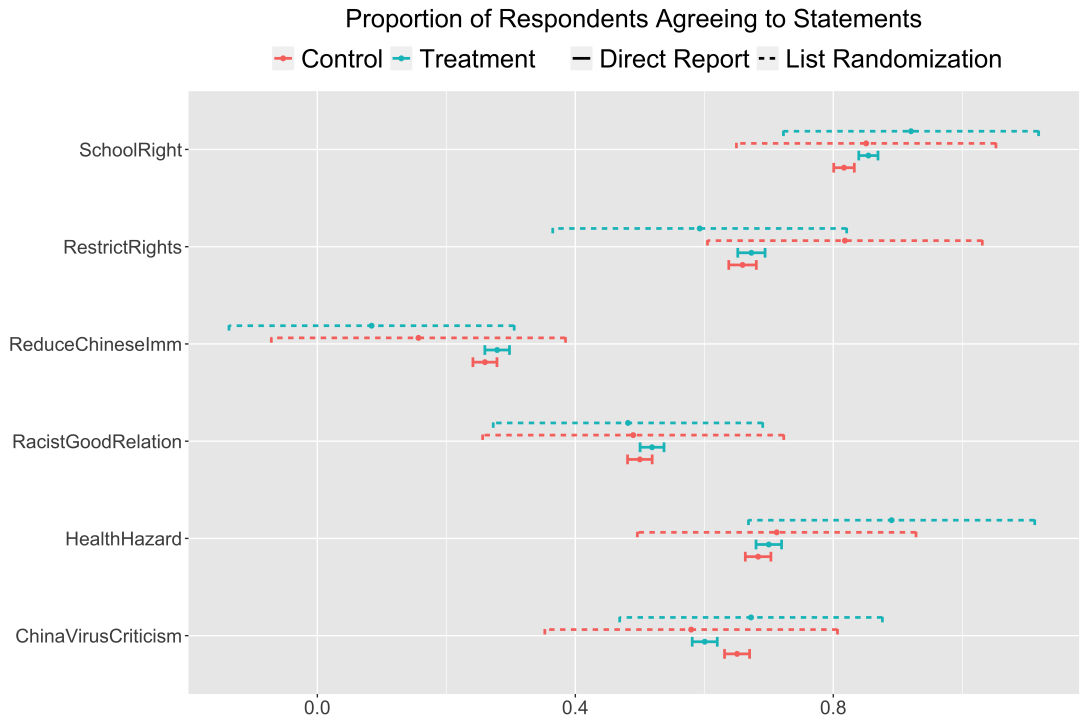


Figure D.1: List Randomization Test by RCT Treatment

Table D.1: Effect of RCT Treatment on Sample Attrition

	Dependent variable:		
	Pass Attn Check 2	State No Bias	Included in Final Sample
RCT Treatment	0.004	0.004	0.006
	(0.015)	(0.013)	(0.015)
R^2	0.00002	0.00004	0.00004
Observations	4,538	2,723	4,538

Note: This table shows the RCT treatment does not change the sample attrition. "Pass Attn Check 2" is whether passing the second attention check screener, "State No Bias" means whether the respondent stated our survey looked neutral, and "Included in Final Sample" means the respondent passed the second attention check screener and stated our survey looked neutral.

E Evidence on Sample Selection

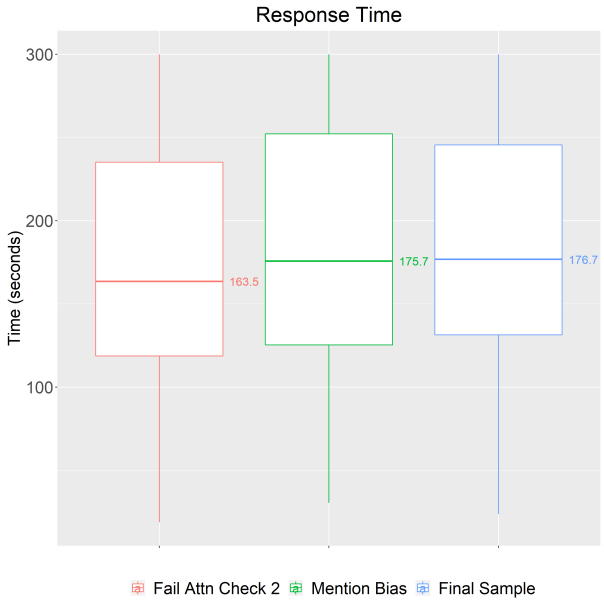
This section presents descriptive statistics about the sample who got screened out for either failure to pass the attention check questions or for their mention about bias in our survey. We found our final sample is slightly different from the dropped sample in multiple dimensions.

Electronic copy available at: https://ssrn.com/abstract=3958536

Importantly, we found evidence that the screened-out sample shows higher racial animus and lower perceived unacceptance, so sample selection makes our final sample less xenophobic. Sample who failed the attention check question spent less time on the response which is shown by smaller 25th, median, and 75th response time in Figure E.1.

Table E.1 shows the means and standard deviations of key demographic variables by screen status. *, **, *** indicates the p-values to test whether the means are different from those of the final sample. Figure E.1 shows the response time in the module asking about racial animus and perceived unacceptance. Figure E.2 shows the average z-score of racial animus and perceived unacceptance. The group who failed the first attention check question is screened out before starting the module on racial animus and perceived unacceptance, so the group is omitted in these graphs.

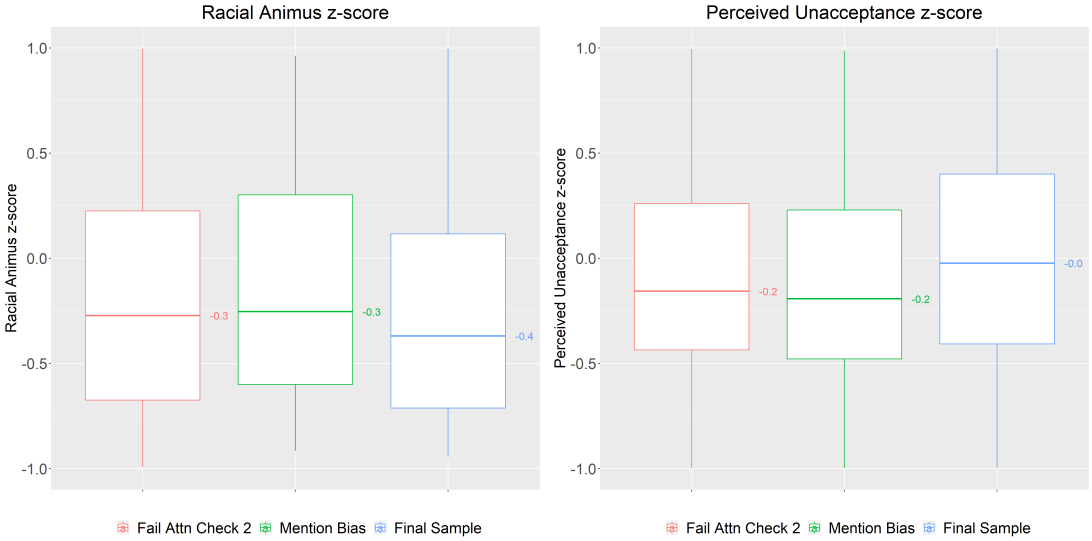
Figure E.1: Response Time for Racial Animus and Perceived Unacceptance Questions



Note: This figure shows the distribution of response time to Racial Animus and Perceived Unacceptance Questions. The median response time is annotated next to each box plot.

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure E.2: Racial Animus and Perceived Unacceptance by Sample Selection



Note: This figure shows the distribution of z-scores of racial animus and perceived acceptance by sample attrition status. The median z-scores are annotated next to each box plot.

Electronic copy available at: https://ssrn.com/abstract=3958536

Table E.1: Descriptive Statistics about Sample Selection

	Fail Attn Check 1	Fail Attn Check 2	Mention Bias	Final Sample
Male	0.43	0.44	0.45	0.45
	(0.5)	(0.5)	(0.5)	(0.5)
18-29 years old	0.22***	0.17*	0.26***	0.19
	(0.41)	(0.37)	(0.44)	(0.39)
30-59 years old	0.52***	0.53***	0.57	0.58
	(0.5)	(0.5)	(0.5)	(0.49)
60-70 years old	0.26*	0.3***	0.17***	0.24
	(0.44)	(0.46)	(0.38)	(0.43)
High School or Below	0.4**	0.35	0.38	0.37
	(0.49)	(0.48)	(0.48)	(0.48)
Some College	0.27	0.27	0.27	0.28
	(0.44)	(0.44)	(0.44)	(0.45)
College	0.34*	0.38	0.36	0.36
	(0.47)	(0.49)	(0.48)	(0.48)
White	0.77***	0.81	0.81	0.8
	(0.42)	(0.39)	(0.39)	(0.4)
Black/African American	0.15***	0.12	0.1	0.11
	(0.36)	(0.32)	(0.3)	(0.31)
Others	0.07***	0.07**	0.09	0.09
	(0.26)	(0.25)	(0.28)	(0.28)
Married	0.5	0.54**	0.49	0.5
	(0.5)	(0.5)	(0.5)	(0.5)
$0~$38754	0.38***	0.31	0.32	0.32
	(0.49)	(0.46)	(0.47)	(0.47)
$38755~$73978	0.25***	0.28**	0.27	0.31
	(0.43)	(0.45)	(0.44)	(0.46)
$73979~$129066	0.21***	0.22	0.27	0.24
	(0.41)	(0.42)	(0.44)	(0.43)
$129067+	0.16**	0.18***	0.14	0.14
	(0.37)	(0.39)	(0.35)	(0.34)
Sample Size	5454	1815	360	2363

Note: This table summarizes how the screened samples are different from our final sample. 'Fail Attn Check 1' and 'Fail Attn Check 2' are the groups who got screened out from the first and second attention check questions. 'Mention Bias' is the group that said our survey looked biased in either direction. 'Final Sample' is the sample that passed all selection criteria and was used for our analysis. The standard deviations are in parentheses. *, **, *** denotes the p-values to test whether the means are different from that of the final sample. * means p<0.10, ** means p<0.05, *** means p<0.01

F Survey Instrument Validation Using Twitter Data

This section describes further details on Twitter data and illustrates how we validate our survey instruments using Twitter-based measures. We asked survey respondents to share their Twitter usernames for research if they have an account. We selected 5 people randomly who

Electronic copy available at: https://ssrn.com/abstract=3958536

Table F.1: Selection in Twitter Sample

	Dependent variable:	
	Have Twitter Account	Share Twitter Account (cond. on having an account)
Racial Animus (Z-score)	−0.015	−0.014
	(0.015)	(0.024)
Perceived Unacceptance (Z-score)	0.065***	0.059**
	(0.018)	(0.029)
Average of Dependent Variable	0.417	0.422
Observations	2,148	887

Note : This table shows that the Twitter sample is selective in the perceived unacceptance of xenophobia.
*p<0.1; **p<0.05; ***p<0.01

shared their Twitter usernames with us and paid them $20 each. 986 participants said they have a Twitter account and among those, 416 survey participants shared a Twitter username with us.

Next, we downloaded the tweets from the usernames that were posted between the beginning of the pandemic, January 1, 2020, and the last survey date, May 24, 2021, and there were 408,116 tweets. Among those tweets, we selected those including keywords related to either coronavirus or Chinese or Asian or words indicating anti-Asian hate.[39] There were 3,727 such tweets. We hired a research assistant to read all 3,727 tweets and to code whether the tweets were anti-Asian, pro-Asian, or irrelevant, and we also read through the labeled tweets again to check for any mistakes. We exclude tweets about Asian foods or music or the Chinese government from either pro-Asian or anti-Asian tweets.

We use Twitter-based measures to validate our other survey instruments and do not use them for structural estimation due to the small sample size and selectivity in image concern. As Table F.1 shows, those who think Sinophobia is unacceptable are more likely to have a Twitter account and are more likely to share their Twitter username with us. Consistent with this, we do not see much hate speech in the Twitter sample: only 1.7% of those who shared

[39]The keywords we used are 'covid', 'coronavirus', 'asia', 'asian', 'beijing', 'ccp', 'china', 'chinese', 'ckmb', 'communist', 'communists', 'cpc', 'huanan', 'hubei', 'jinping', 'patient zero', 'prc', 'tedros', 'wuhan', 'xi jinping', 'xijinping', 'xinnie', 'aseng', 'bamboo coon', 'bamboo coons', 'bat eater', 'bioterrorism', 'bioweapon', 'boycottchina', 'ccpvirus', 'chinadidthis', 'chinaliedpeopledie', 'chinaliedpeopledied', 'chinaman', 'chinamen', 'chinavirus', 'ching chong', 'chink','chinks' , 'chinky', 'cokin', 'commie', 'commies', 'communistvirus', 'coolie', 'dog eater', 'fuckchina', 'kungflu', 'ling ling', 'makechinapay', 'niakoué', 'pastel de flango', 'sideways cooters', 'sideways pussies', 'sideways pussy', 'sideways vagina', 'sideways vagina', 'slant-eye', 'slopehead', 'ting tong', 'wuflu', 'wuhanflu', 'wuhanvirus', 'kung flu'.

Electronic copy available at: https://ssrn.com/abstract=3958536

Table F.2: Correlation between Pro- and Anti-Asian Tweets and Survey Instruments

	Dependent variable:					
	Any Pro-Asian Tweet			Any Anti-Asian Tweet		
Racial Animus (Z-score)	−0.046***		−0.028	−0.001		−0.008
	(0.016)		(0.019)	(0.008)		(0.010)
Perceived Unacceptance (Z-score)		0.074***	0.058**		−0.012	−0.016
		(0.020)	(0.023)		(0.010)	(0.012)
Average of Dependent Variable	0.07	0.07	0.07	0.017	0.017	0.017
Observations	413	385	384	413	385	384

Note: This table shows the correlation between pro- and anti-Asian tweets and our survey instruments on racial animus and perceived unacceptance. *p<0.1; **p<0.05; ***p<0.01

Table F.3: Correlation between Xenophobic Action Survey Measures and Pro-, Anti-Asian Tweets

	Dependent variable:		
	Xenophobic Donation	Xenophobic Petition	(DG)1(White>Chinese)
Any Pro-Asian Tweet	−0.101	−0.097*	−0.049
	(0.071)	(0.055)	(0.053)
Any Anti-Asian Tweet	−0.013	0.062	−0.079
	(0.141)	(0.109)	(0.105)
Average of Dependent Variable	0.163	0.089	0.082
Observations	416	416	416

Note : This table shows the correlation between pro- and anti-Asian tweets and our survey instruments on xenophobic action. *p<0.1; **p<0.05; ***p<0.01

a Twitter username posted any anti-Asian tweet during the pandemic, whereas 7% of those posted any pro-Asian tweet during the pandemic. It is possible that people who posted many hate speech tweets might be more reluctant to share their Twitter usernames with us.

The Twitter-based measures validate that our survey instruments are not cheap talk. We

Electronic copy available at: https://ssrn.com/abstract=3958536

Table F.4: Correlation between Pro- and Anti-Chinese Tweets and Survey Instruments

	Dependent variable:					
	Any Pro-Chinese Tweet			Any Anti-Chinese Tweet		
Racial Animus (Z-score)	−0.029***		−0.016	−0.001		−0.008
	(0.011)		(0.013)	(0.008)		(0.010)
Perceived Unacceptance (Z-score)		0.048***	0.039**		−0.012	−0.016
		(0.014)	(0.016)		(0.010)	(0.012)
Average of Dependent Variable	0.031	0.031	0.031	0.017	0.017	0.017
Observations	413	385	384	413	385	384

Note: This table shows the correlation between pro- and anti-Chinese tweets and our survey instruments on racial animus and perceived unacceptance. *p<0.1; **p<0.05; ***p<0.01

find survey instruments on racial animus and perceived unacceptance are strongly related to whether a person posted any pro-Asian tweet during the pandemic (Table F.2). We do not find such a correlation for the anti-Asian tweet, however, potentially due to selection in our sample and scarcity of such tweets. Table F.3 shows a correlation between Twitter measures and other xenophobic behavior measures. Pro-Asian tweets are negatively correlated with other xenophobic behavior measures: such correlation is significant for the xenophobic petition at the 10% significance level. The correlation with the xenophobic donation is weaker than that: it is significant at about 16% significance level. It is important to consider the coefficients' size as well as statistical significance: the coefficient sizes of all action measures are large, between 62% and 108% of the means of dependent variables. Not surprisingly, we do not find evidence of correlation with the anti-Asian tweet, which is consistent with Table F.2.

One may be concerned about using pro- and anti-Asian tweet measures instead of narrowly defined pro- and anti-Chinese tweet measures. We chose to use pro- and anti-Asian tweet measures instead because social movements on Twitter in response to the rise of xenophobia used the pro-Asian phrases more often, as can be seen from popular hashtags, such as "♯StopAsianHate" or "♯StopAAPIHate". However, we replicated our analysis using narrowly defined pro- and anti-Chinese tweet measures for robustness check. The results are similar to our baseline results, but the correlations between some of our survey instruments and the pro- and anti-Chinese tweet measures become slightly weaker (Table F.4, F.5).

Electronic copy available at: https://ssrn.com/abstract=3958536

Table F.5: Correlation between Xenophobic Action Survey Measures and Pro-, Anti-Chinese Tweets

	Dependent variable:		
	Xenophobic Donation	Xenophobic Petition	(DG)1(White>Chinese)
Any Pro-Chinese Tweet	−0.169	−0.096	−0.080
	(0.105)	(0.081)	(0.078)
Any Anti-Chinese Tweet	−0.002	0.066	−0.074
	(0.142)	(0.109)	(0.105)
Average of Dependent Variable	0.163	0.089	0.082
Observations	416	416	416

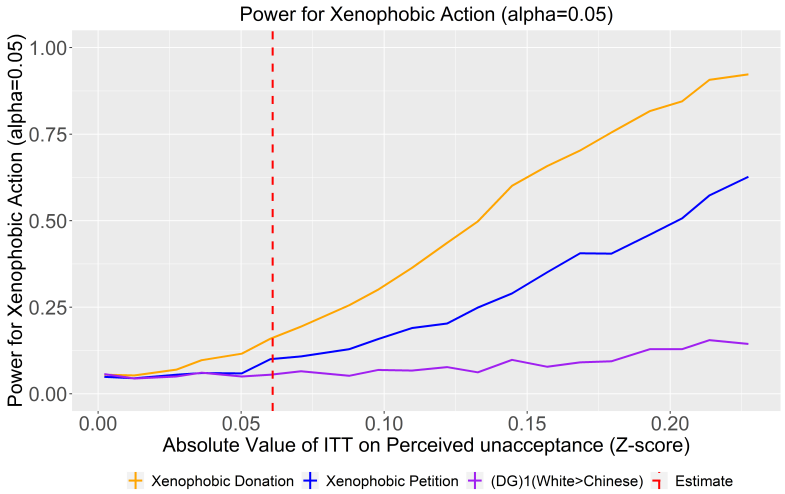
Note : This table shows the correlation between pro- and anti-Chinese tweets and our survey instruments on xenophobic action. *p<0.1; **p<0.05; ***p<0.01

G Monte Carlo Evidence on the Power of Testing the Treatment Effect

We provide Monte Carlo evidence on the probability of finding significant treatment effects on xenophobic actions conditional on having a significant treatment effect on the perceived unacceptance z-score. To do this, we simulate data of the same sample size, 2363 individuals, and randomly assign half of the simulated sample to a treatment group. Next, we draw racial animus and perceived unacceptance for the simulated sample using the estimated distribution of racial animus and perceived unacceptance, but for the treated group, we use the shifted perceived unacceptance distribution. This is consistent with the evidence in Table 5 that the information RCT treatment effect on the perceived unacceptance is significant under 5% significance level but is insignificant for the racial animus. We vary the degree of the shifts in perceived unacceptance distribution. After drawing racial animus and perceived unacceptance, we simulate proxies of racial animus, perceived unacceptance, and xenophobic actions using the structural parameter estimates. Finally, we run the same RCT regressions in Table 5 and D.1 and count how many times out of 1000 times we find the significant treatment effect on xenophobic actions under the 5% significance level. We also compute the average of treatment effect estimates on the perceived unacceptance z-score.

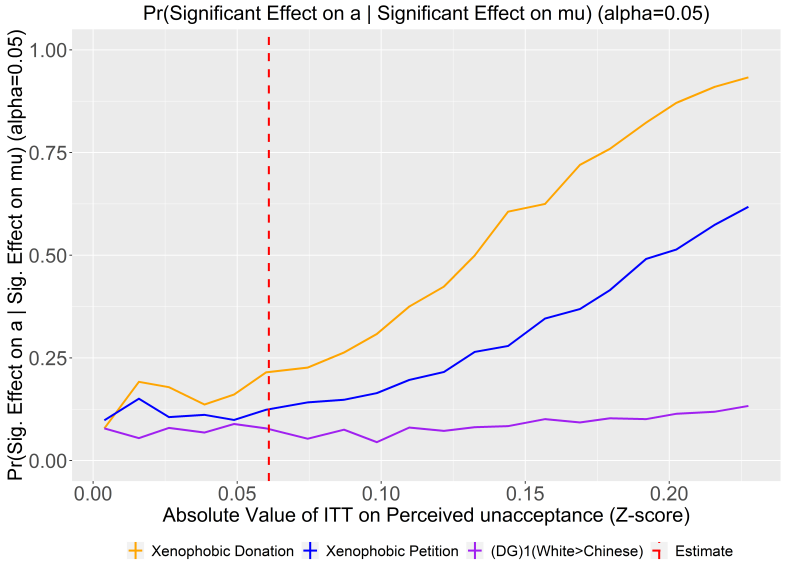
Electronic copy available at: https://ssrn.com/abstract=3958536

Figure G.2: Power Analysis for Treatment Effects on Actions



Note: This figure shows the power of testing a treatment effect on xenophobic actions when we shift the distribution of perceived unacceptance to different degrees. For each degree of shift in perceived unacceptance distribution, we simulate a sample of the same size with our data, 2363 individuals, 1000 times. The red bar shows the estimate of the treatment effect on perceived unacceptance z-score (0.61, Table 5). Under this estimate, the powers for xenophobic actions are between 6% and 16%.

Figure G.1: Probability to Find a Significant Treatment Effect on Actions Conditional on Having a Significant Treatment Effect on Perceived Unacceptance Z-score

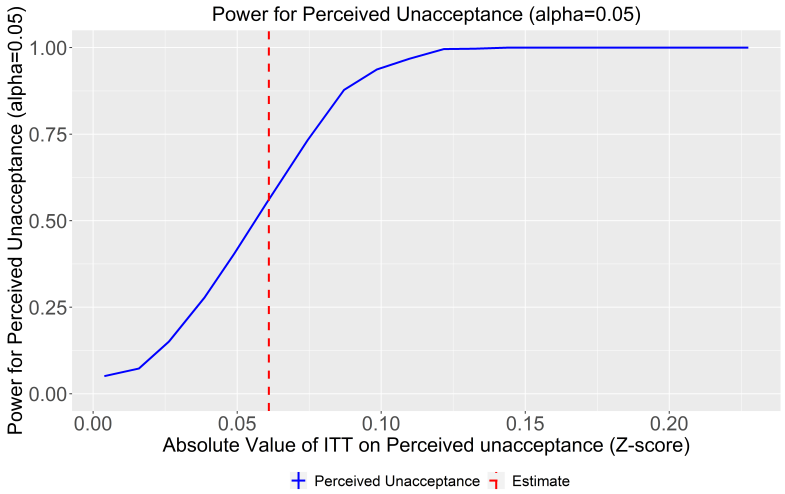


Note: This figure shows the probability to find a significant treatment effect on actions (a) conditional on having a significant treatment effect on perceived unacceptance z-score (Mu). For each degree of shift in perceived unacceptance distribution, we simulate a sample of the same size with our data, 2363 individuals, 1000 times. The red bar shows the estimate of the treatment effect on perceived unacceptance z-score (0.61, Table 5). Under this estimate, probabilities are between 8% and 21%.

Figure G.1 illustrates the main takeaway from this exercise. Since we vary the degree of shift in the distribution of perceived unacceptance, the absolute value of the ITT estimates on the perceived unacceptance z-score also varies. The red vertical bar shows the current

Electronic copy available at: https://ssrn.com/abstract=3958536

Figure G.3: Power Analysis for Treatment Effect on Perceived Unacceptance



Note: This figure shows the power of testing a treatment effect on perceived unacceptance when we shift the distribution of perceived unacceptance to different degrees. For each degree of shift in perceived unacceptance distribution, we simulate a sample of the same size with our data, 2363 individuals, 1000 times. The red bar shows the estimate of the treatment effect on perceived unacceptance z-score (0.61, Table 5).

treatment effect estimate on the perceived unacceptance z-score. Under this estimate, the probabilities of finding a significant treatment effect on actions are between 8% and 21%. If the treatment effect on perceived unacceptance z-score were stronger, say three times the current estimate, then the probabilities of finding a significant treatment effect on xenophobic actions conditional on having the significant treatment effect on the perceived unacceptance are between 10% and 76%. The probability is biggest for a xenophobic donation, 76%, whose κ parameter is the largest, and the probability is smallest for a dictator game outcome, 10%, whose κ parameter is the smallest. This result is intuitive because in the discrete choice model in equation 1, μ is multiplied with κ. So if κ is larger, the treatment effect on actions is more likely to be significant when the μ distribution is shifted.

Figure G.2 shows that the probabilities of finding a significant treatment effect on xenophobic actions (unconditionally) are all small – less than 16% – and similar to Figure G.1. If the treatment effect on perceived unacceptance were larger – say, three times the current estimate – the powers of testing the treatment effect on xenophobic actions becomes much bigger – between 9% and 76%.

Figure G.3 shows the probability of finding the significant treatment effect on the perceived unacceptance. The probability is much higher than the one for the xenophobic actions under the current estimates, that is, 55%.

Electronic copy available at: https://ssrn.com/abstract=3958536