ESSAYS ON HATE CRIME, MEDIA, AND CORRUPTION

A Dissertation

by

ANDY CAO

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Danila Serra
Committee Members,	Jennifer Doleac
	Catherine Eckel
	Marco Palma
Head of Department,	Steve Puller

May 2023

Major Subject: Economics

Copyright 2023 Andy Cao

ABSTRACT

My dissertation consists of three essays on the topics of hate crime, media, and corruption. In the first chapter, I investigate the role news media plays in promoting hatred through the news coverage of mass shootings. I first show through observational data that the media treats hate-motivated mass shootings differently by focusing more on the shooter, and this possibly results in an increase in hate crimes against the group that was targeted in the shooting. I then design and run an online information experiment to causally examine the impact of news coverage on spreading hatred. Results from the experiment show that receiving details about the shooter's hate ideology increases Republicans' support for the shooter. Emphasis on the shooter's identity and background increases Democrats' support for both the shooter and the shooter's hate ideology. The latter finding is driven by the more right-leaning individuals within the Democrat sample. My findings highlight media's role in spreading hatred and provide important guidelines on media's approach to hate-motivated mass shootings.

In the second chapter, coauthored with Jason Lindo and Jiee Zhong, we study the influence of high-profile individuals on anti-social behaviors. In particular, we investigate whether Donald Trump's "Chinese Virus" tweets contributed to the rise of anti-Asian incidents. We find that the number of incidents spiked following Trump's initial "Chinese Virus" tweets and the subsequent internet search activity for the phrase rose dramatically. Difference-in-differences and event-study analyses indicate that this spike was significantly more pronounced in counties that supported Donald Trump in the 2016 presidential election. Our study shows that high-profile individuals such as Trump can have detrimental effects, even when the technology of social media substantially limits what they can say. Our findings have important implications given the recent rise of populist leaders pushing anti-social beliefs and behaviors on topics ranging from vaccine hesitancy to the treatment of immigrants.

In the third chapter, coauthored with Dmitry Ryvkin and Danila Serra, we use a laboratory experiment to study the self-selection of individuals into committees that have discretionary power over the distribution of public resources. We examine how the status quo level of corruption and the individual's propensity for corruption affect the decision to join the committee. Results from our experiment show that subjects have higher interest in joining a corrupt committee compared to an honest committee, regardless of their own propensity for corruption. We also find evidence that committee members' voting behavior and communication pattern appear in line with their "type", i.e., corrupt individuals are more likely to support embezzlement. Taken altogether, our results highlight the importance of the screening process for public servants. A screening method that focuses on characteristics such as honesty and pro-sociality can be an effective way to reduce corruption.

DEDICATION

Rose is a rose is a rose is a rose.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Danila Serra for her guidance and support. Having her as my advisor is the best decision I have made in the Ph.D. program. I would like to thank Jennifer Doleac, Catherine Eckel, Marco Palma, as well as other faculty members of the Department of Economics for their invaluable feedback on my research. I would also like to thank Jason Lindo, Dmitry Ryvkin, and Jiee Zhong for being the best coauthors.

I am grateful to my parents and my sister for their support, two of whom are PhDs themselves. Yet, they resisted the temptation to advise me on research and let me study the topics that I am passionate about.

I am very fortunate to have married my wife, Xinyi Wang, on July 10th, 2022. Although her masculinity makes me question my gender identity at times, the thoughts of one day graduating, ending long-distance, and starting a family with her have kept me going forward.

I am happy that I have made some friends in College Station. In particular, I would like to thank Sam Priestley for building up my alcohol tolerance, Daniel Gomez for spicing everything up, and the habibis of my cohort, Maya Mikdash and Sora Youn for all the wild discussions we had.

Finally, I would like to thank Charles H. Gregory, whose generous donation made the first chapter of this dissertation possible.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Danila Serra (Chair), Professor Jennifer Doleac, and Professor Catherine Eckel of the Department of Economics, and Professor Marco Palma of the Department of Agricultural Economics.

Chapter II is a joint work with Jason Lindo and Jiee Zhong of the Department of Economics. Chapter III is a joint work with Dmitry Ryvkin of Florida State University and Danila Serra.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Chapter I was supported by research funding from the Department of Economics, and the Charles H. Gregory Class of ‘64 Graduate Strategic Fellowship from the College of Liberal Arts (now College of Arts and Sciences). Chapter III was supported by the John Van Huyck Endowed Fellowship from the Department of Economics.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. DOES NEWS COVERAGE OF HATE-MOTIVATED MASS SHOOTINGS GENERATE MORE HATRED?

1.1 Introduction

[1] Hate crimes have risen to their highest level in over a decade in the United States and have been elevated to top national threat priority.[2] The most extreme and dangerous hate crime is when the offender conducts a mass shooting targeting victims of a specific ethnicity, race, gender, sexual orientation, or religion. I call such act a hate-motivated mass shooting. Hate-motivated mass shootings receive great attention in the news and on social media. For instance, the 2022 Buffalo shooting[3] was headlined by the Wall Street Journal headline for four days straight after the shooting took place. During the same time, it was covered in Television news 767 times and was the most trending topic on Twitter.[4] A number of studies in criminology and social psychology suggest that media coverage of mass shootings could inspire more mass shootings through behavioral contagion (for example, see Meindl & Ivy, 2017 and Langman, 2018). Relatedly, experimental investigations in the laboratory show that individuals are more likely to engage in anti-social behaviors when they get information about others engaging in such behavior (for example, see Gino et al., 2009 and Dimant, 2019). This suggests that media coverage of hate-motivated mass shootings, by focusing on the shooter's ideology and background, may induce or encourage others to express support for such ideology, leading to more hatred toward the victimized group.

In this paper, I employ quasi-experimental and experimental methods to address the following

[1]I would like to especially thank my advisor Danila Serra for her guidance and support. I thank Elliott Ash, Andrew Barr, Alex Brown, Leonardo Bursztyn, Marco Castillo, Jennifer Doleac, Kwabena Donker, Catherine Eckel, Russell Golman, Joanna Lahey, Judd Kessler, Silvana Krasteva, Jonathan Meer, Marco Palma, Ragan Petrie, Santiago Saavedra, Silvia Saccardo, Tim Salmon, Erik Snowberg, and numerous seminar participants for helpful comments. I am grateful for funding support from the Department of Economics and the College of Liberal Arts at Texas A&M University. This study is approved by the Institutional Review Boards at TAMU (IRB2021-1000M). The online experiment is pre-registered on AsPredicted (#74996).

[2]See FBI's release, June 2021, and October 2021.

[3]The shooting took place On May 14th, 2022, in Buffalo, New York. 10 Black people were killed.

[4]WSJ's digital archive of top headlines can be accessed here. The Buffalo shooting was headlined from May 15th to May 18th. Historical television news coverage can be found on the Internet Archive. Historical Twitter Trends can be found on Trend Calendar. The 2022 Buffalo shooting was the number 1 trending topic on May 15th, 2022.

questions. First, are hate-motivated shootings covered differently by the media in terms of intensity and content of coverage? Second, is there more public interest in hate-motivated shootings as compared to non-hate-motivated shootings, everything else being equal? Third, does emphasizing the ideology and/or the identity of the shooter in the coverage of hate-motivated mass shootings affect individuals' attitudes toward the shooter and/or the shooter's ideology, possibly leading to more hatred?

I categorize every notable mass shootings in the United States into hate-motivated mass shooting and non-hate-motivated mass shooting based on the shooter's motive and the characteristics of the victims. Using data from the Internet Archive and the Vanderbilt Television News Archive, I compile a dataset of mass shooting news coverage. I find that the media gives differential treatment to mass shootings based on the nature of the shooting. Whenever a mass shooting is hate-motivated, it receives more news coverage, as measured by the number of news clips and the duration of news clips. Hate-motivated mass shootings on average receive more than three times the amount of coverage on national television networks. This difference is robust even after controlling for the number of casualties and a rich set of fixed effects. I then turn to examine the content of media coverage. For each mass shooting in my dataset, I scrape down news articles published within 14 days after the shooting. I use Natural Language Processing to identify whether a news article is about the shooter. I show that whenever a mass shooting is hate-motivated, news articles are 28.2 percent more likely to feature the shooter. Moreover, I show that a large proportion of these articles are related to the shooter's ideology/motive.

Next, I investigate whether viewers exhibit similar bias in their news preference. I use Google Trends data to examine the online searching behavior related to each mass shooting. I show that people's search interest value increases by 167 percent when a mass shooting is hate-motivated. A closer look at the searched terms suggests this gap is likely driven by the interest in the shooter. For more than 60 percent of the hate-motivated mass shootings, information related to the shooter is in the list of the most searched topics. In many cases, people directly used the name of the shooter and searched for the shooter's manifesto. This pattern exactly matches media's tendency to focus on the shooter when reporting hate-motivated mass shootings. One concern is that the prolonged exposure to the shooter could cause people to copy the shooter's beliefs and behavior, i.e., become hateful toward the population targeted by the shooter, or in the most extreme case, commit hate crimes. To assess this possibility, I use the FBI's hate crime data to compare the number of hate crimes targeting the same population before and after a hate-motivated mass shooting. Using an event study framework, I show that immediately following a hate-motivated mass shooting, there is an increase in the number of hate crimes against the same victimized group in the shooting. In contrast, there is no change in the number of hate crimes against other populations.

Based on my findings and the existing literature, I hypothesize that media coverage of hate-motivated mass shootings generates more support for the shooter and the shooter's hate ideology, possibly leading to more hatred. However, it is challenging to establish causality using observational data. Indeed, even if there is a correlation between media consumption and hatred, that could simply be due to more hate-prone individuals self-selecting into watching or reading news that cover details of hate-motivated mass shootings, rather than to news coverage persuading individuals to be more hateful.

To overcome identification challenges, I employ an online information provision experiment (see Haaland et al., 2020 for a review). I ask subjects to read a piece of news story about the 2019 El Paso shooting, which killed 23 people, including 8 Mexicans. The shooter posted a manifesto online with white nationalist and anti-immigrant themes minutes before the shooting. Subjects are randomly assigned to one of the four treatment conditions which vary in the level of informativeness. The No Hate treatment represents a streamlined version of news coverage. The news story only includes basic description of the shooting, such as the location and the number of victims. The Hate treatment is identical to the baseline treatment, except that the news story now mentions the shooting was targeted at the local Hispanic community and is possibly a hate crime. The Hate Ideology treatment builds on the Hate treatment and adds extra information on the shooter's ideology, i.e., why he targeted the Hispanic population. Finally, the Hate Background treatment builds on the Hate Treatment and provides additional information on the shooter's background, including name, photo, and quotations from former classmates hinting the shooter had a difficult childhood and was bullied at school by Hispanic students. My outcome variables are measures of: 1) Interest in the shooting, measured by whether subjects ask to be shown more information at the end of the survey; 2) Attitudes toward the shooter, measured by survey questions on admiration for the shooter, justification for the shooter's action, and sentencing option for the shooter; 3) Attitudes toward the ideology of the shooter, measured by an $1 donation to an anti-immigrant organization or a pro-immigrant organization following the methodology used by Bursztyn, Egorov, and Fiorin (2020); 4) Hatred, measured by interest in accessing information on a white supremacy hate group.

Between Fall 2021 and Fall 2022, I recruited 2,400 American men to participate in the study via the online platform Prolific and CloudResearch. I stratified the recruitment and treatment assignment by political affiliation, with the aim of including an equal number of Republicans and Democrats in my sample. My final sample consists of 1199 Democrats and 1201 Republicans. This design choice allows me to test for possible heterogeneity in the impact of news reporting on individual attitudes, since the two parties have very polarized views regarding immigrants (Card et al., 2022).

I have four main findings from the experiment. My first finding is that subjects are not intrinsically more interested in hate-motivated than non-hate-motivated shootings. There is no significant difference in information demand between the No Hate treatment and the Hate treatment. This is true both for the Republican and the Democrat samples. This suggests that the higher public interest in hate-motivated mass shootings I observe in the online searching data is likely caused by either the higher intensity in media coverage of hate-motivated mass shootings as compared to non-hate motivated mass shootings, or in the way the two types of shootings are covered (i.e., the content of coverage). However, when subjects read the more informative news stories in the Hate Ideology treatment and the Hate Ideology Background treatment, they show significantly less demand for information on the shooting (both on the shooter and his ideology), than when such information is not provided, in the Hate treatment. Therefore, the higher public interest observed in the Google search data is more likely to be driven by the higher intensity of media coverage of hate-motivated shootings, rather than the differences in the content of media coverage.

My second finding is that the way the shooting is covered in the news affects subjects' attitudes toward the shooter. To start with, both Democrat and Republican subjects decrease their support for the shooter when they know that the shooting is hate-motivated. They believe the shooter deserves a higher sentence, is less admirable, and his action less justifiable. This suggests that people have a natural distaste for hate crimes. However, the decrease in Republican subjects' support goes away when they are provided additional information on the shooter's ideology. I construct a standardized index of support for the shooter (Anderson, 2008) and show that Republican subjects in the Hate Ideology treatment increase their support for the shooter by 0.19 standard deviations relative to the Hate treatment (p-value=0.005). This suggests when Republican subjects are given details explaining the shooter's ideas and beliefs about why he targets immigrants, they become more approving. In contrast, while knowing the shooter's ideology does not affect Democrat subjects' attitudes toward the shooter, knowing the shooter's background story about his childhood suffering increases their support for the shooter by 0.27 standard deviations (p-value=0.00).

My third finding is that when exposed to the news story that emphasizes the shooter's background, Democrat subjects significantly increase their support for the shooter's anti-immigrant ideology, as measured by donations to an anti-immigrant organization. In the experiment, about 75% of the subjects are randomly assigned to an anti-immigrant organization, while the rest of the 25% are assigned to a pro-immigrant organization. In both cases, near the end of the survey, subjects are given the description of the organization and asked whether they want to authorize a $1 donation to the randomly assigned organization.[5] On average, the donation rate to the pro-immigrant organization is significantly higher in all treatments among both Republicans and Democrats. As expected, the donation rate to the anti-immigrant organization is significantly higher among Republicans (23.21 percent versus 14.49 percent, p-value=0.00). However, consistent with the increase in support for the shooter, Democrat subjects in the Hate Background treatment are 6.9 percentage points more likely to authorize a $1 donation to the anti-immigrant organization relative to

[5]The anti-immigrant organization is the Federation for American Immigration Reform. The pro-immigrant organization is the American Immigration Council.

a mean of 11.4% in the Hate Treatment (p-value=0.037). In contrast, Republican subjects in the Hate Background treatment are 9.3 percentage points less likely to donate to the anti-immigrant organization relative to a mean of 27.8% in the Hate Treatment (p-value=0.022). This is surprising, given that Democrat subjects are generally known for their friendly attitudes toward immigrants. Further analysis shows that the treatment effect on Democrat subjects are driven by the more right-leaning individuals within the Democrat sample. There is no change in the donation rate to the pro-immigrant organization across different treatment conditions from either Democrat subjects or Republican subjects.

Finally, I provide suggestive evidence that exposure to the shooter's background story increases Democrat subjects' interest in white supremacy hate groups. At the end of the survey, subjects are given a brief description of a major hate group known for its white nationalist and white supremacy themes, and are told that this group shares similar ideology with the shooter. Subjects are then asked: "Would you like to know how to access its website". I track whether subjects click on the links, which are provided if and only if subjects answered "yes" in the previous question. Results show that only 8.51 percent of Democrat subjects and 12.82 percent of Republican subjects requested the links (p-value=0.00), and even fewer subjects clicked on the links. Consistent with the increase in support for the shooter and the shooter's ideology, Democrat subjects in the Hate Background treatment are 3.95 percentage points more likely to request the links to the hate group's website. This difference is statistically significant (46.47 percent increase, p-value=0.076), despite the small sample size. However, knowing the shooter's background story does not make Democrat subjects more likely to click on the links. While there is some evidence that Republican subjects in the Hate Ideology Treatment increases their likelihood of clicking on the links, the difference is not statistically significant (61.32 percent increase, p-value=0.265)

In summary, my experiment finds substantial variation in treatment effects based on subjects' baseline political stance. At the aggregate level, Republican subjects show higher support for the shooter, donate more to the anti-immigrant organization, and express higher interest in the white-supremacy hate group. A news story that emphasizes the shooter's anti-immigrant ideology increases Republican subjects' support for the shooter. A news story that emphasizes the shooter's identity and background, and highlights the shooter's struggles growing up increases Democrat subjects' support for the shooter, support for the shooter's ideology (more than 60 percent increase in donation rate to an anti-immigrant organization), and interest in white supremacy hate groups (suggestive evidence). Taken altogether, the heterogeneity in treatment effects suggests that Republican subjects respond to the shooter's ideology itself, while Democrat subjects respond to the shooter's traumatic background story which hints at the shooter being bullied by Hispanic students and possibly explains how the shooter developed his ideology.

My paper contributes to the literature on media coverage of mass shootings and its unintended consequences (see Lankford & Madfis, 2018a for a review). There has been a long-standing debate on how media should report mass shootings. Many studies point out that future offenders often find inspirations from past shooters (Helfgott, 2015; Kissner, 2016; Lankford, 2016; Meindl & Ivy, 2017; Murray, 2017; Langman, 2018; Lee, 2018). Some shooters even personally acknowledged that they were influenced and motivated by past mass shooters.[6] This suggests that media reporting on mass shootings, i.e., the victims, the suspect and his/her motives, may play a role in creating similar crimes. However, most research on this topic relies on correlational analysis and lacks proof of causality.[7] One important exception is Jetter and Walker (2022), who use exogenous variations in worldwide disaster deaths to show that news coverage of mass shootings causes more subsequent mass shootings. I extend Jetter and Walker (2022) by focusing on the effects of the content of mass shooting news coverage, rather than the intensity, and by focusing specifically on hate-motivated mass shootings. My experiment exogenously varies the information provided in the news story about a mass shooting to test how different kinds of information affect individuals' reactions to the crime, and in particular attitudes toward victims vis-à-vis the shooter. To the best of my knowledge, this is the first paper to causally show that media coverage of mass shootings can increase support for the shooter and the shooter's hateful ideology. The results from my experiment

[6] For instance, at least 32 attackers referred to the 1999 Columbine shooters as role models (Langman, 2018).

[7] Though researchers have proposed regulations such as calling for the media to stop publishing information about mass killers (for example, see this open letter signed by 149 scholars and professionals), such policies have never been tested and are largely ignored in practice.

have important implications for media and policy makers.

More generally, my paper contributes to the vast literature that studies the influence of media on a variety of outcomes, including voting behavior and political opinions (DellaVigna & Kaplan, 2007; Gerber et al., 2009; Martin & Yurukoglu, 2017), immigration (Couttenier et al., 2021; Schneider-Strawczynski & Valette, 2021), terrorism (Jetter, 2017; Durante & Zhuravskaya, 2018), criminal justice (Lim et al., 2015; Mastrorocco & Minale, 2018; Philippe & Ouss, 2018), socio-economic development (La Ferrara, 2016), and precautionary behavior during the Covid-19 pandemic (Bursztyn, Rao, et al., 2020; Simonov et al., 2020). However, there is little evidence of media's ability to promote hatred. The exceptions are Yanagizawa-Drott (2014) and Adena et al. (2015), who examines the impact of government propaganda on mass violence, Ang (2020) who examines the impact of blockbuster film on racism, Bursztyn et al. (2019), Müller and Schwarz (2020), and Carr et al. (2020), who examines the impact of social media on racism. In previous work (Cao et al., 2022), I show that Donald Trump's "Chinese Virus" tweets contributed to the rise of anti-Asian incidents. My paper focuses on news media, which represents a very different context. On the one hand, news media is (supposed to be) neutral and objective in nature. On the other hand, news media is people's primary information source for criminal incidents and is therefore, especially hard to regulate. To the best of my knowledge, my paper is the first to document media's differential treatment of mass shootings based on the underlying motive of the shooting. My experiment shows that the way hate-motivated mass shootings are covered in the news could lead to unintended consequences including more support for the shooter and the shooter's hateful ideology. Thus, my paper extends the literature on media influence by showing that news media is capable of causing extreme behaviors.

Finally, my paper relates to the growing body of literature that uses information provision experiments (for a review, see Haaland et al., 2020). In particular, this paper adds to the studies on attitudes related to immigrants and immigration. Alesina et al. (2018) finds that providing true information about shares and origins of immigrants does not increase self-reported support for redistribution and donation to charity. Similarly, Grigorieff et al. (2020) finds that factual information about the characteristics of immigrants does not affect policy preferences measured by donation and petition. Bursztyn, Egorov, and Fiorin (2020) finds that information about Trump's popularity increases individuals' willingness to publicly express xenophobic views as measured by donation to an anti-immigrant organization. Haaland and Roth (2020) finds that providing subjects with research evidence on the labor market impacts of immigration makes subjects more supportive of immigration measured by self-reported views and petition signatures. My paper deviates from these studies in two ways. First, my information treatments resembles news articles that people may encounter regularly. I show that even a small variation in the news' content about a mass shooting could affect how people feel towards immigrants. My findings suggest that narratives might be more effective in changing people's attitudes than facts (Bénabou et al., 2018). Second, my outcome measures aim at capturing extreme behaviors such as support for the shooter and hatred, rather than strong support for immigration control. My paper shows that information experiments can be applied to study anti-social behaviors.

Overall, my paper highlights media's role in spreading hatred. My findings provide important guidelines on media's approach to hate-motivated mass shootings, and more broadly, crimes. To reduce the risk of hate crimes, it is desirable not to provide detailed information on shooter's ideology, as well as the shooter's identity and details about the shooter's background that focus on his troubled upbringing. The remainder of this paper is structured as follows. Section 1.2 presents evidence from observational data. Section 1.3 introduces my hypotheses and experimental design. Section 1.4 presents results from the experiment. Section 1.5 concludes.

1.2 Evidence from Observational Data

This section contains my motivational analysis using observational data. Section 1.2.1 discusses how I categorize mass shootings in the US into hate-motivated and non-hate-motivated. Section 1.2.2 documents the difference in news coverage of hate-motivated mass shootings and non-hate-motivated mass shootings. Section 1.2.3 shows the difference in viewers' reactions. Section 1.2.4 provides evidence that following a hate-motivated mass shooting, there is an increase in the number of hate crimes against the same group that was targeted in the shooting.

1.2.1 Mass Shootings in the US

I collect data for mass shootings in the United States from the list of the most notable mass shootings in the United States complied by Wikipedia.[8] Wikipedia defines a mass shooting as 4 or more shot in one incident.[9] The dataset contains 241 mass shootings in the United States from 1970 to March 2021.[10] I choose Wikipedia over other data sources for two reasons. First, it consists of mass shootings that attracted wide media coverage and public discussion. Second, these shootings tend to have a more complete description of the incident including the shooter's motive. These two features help me address my research questions.

For the purpose of this study, I divide notable mass shootings in the United States into two categories: hate-motivated mass shootings and other mass shootings. Based on the FBI's definition of a hate crime, I define a hate-motivated mass shooting as a mass shooting which is motivated, in whole or in part, by the offender's bias(es) against a: race, religion, sexual orientation, ethnicity, gender, gender identity. For each mass shooting in my dataset, I classify it as hate-motivated if it meets one of the following conditions:

1. The suspect explicitly admitted that the shooting was motivated by hatred/bias.
2. The majority of the victims belong to the same group such that the authorities were investigating the possibility of a hate crime against that group.

Note that if it is the latter case, there is a chance that the investigation failed to prove the shooting was a hate crime. Therefore, I could incorrectly label a mass shooting as hate-motivated when in reality it is not. However, since the shooting was investigated as a potential hate crime, being reported by the media as a potential hate crime, generated public discussion about hate crimes, it should still be classified as a hate-motivated mass shooting.

[8] Only shootings that have Wikipedia articles of their own are included in this list.

[9] Note that there is no agreement on how mass shooting is defined. Some agencies use a more restrictive definition, for example, Mother Jones defines a mass shooting as three or more shot and killed in one incident at a public place, excluding the perpetrators.

[10] This list is more inclusive compared to alternative sources such as Mother Jones (120 Mass shootings), Washington Post (180 Mass shootings), and The Violence Project (170 Mass shootings).

Based on my definition, 24 out of 241 (10%) mass shootings are identified as hate-motivated. A complete list of hate-motivated mass shootings is provided in Table A.1 in the Appendix. Appendix Table A.2 shows the descriptive statistics. On average, hate-motivated mass shootings are associated with slightly higher casualty. However, the difference is not statistically significant (the p-value of a t-test of equality is 0.13 and 0.67 respectively for the difference in mean of dead and injured).

1.2.2 News Coverage of Mass Shootings in the US

In this section, I compile and analyze data on media coverage of the most notable mass shootings in the United States. I provide evidence that, compared to non-hate-motivated shootings, the media coverage of hate-motivated mass shootings: 1) lasts longer, in terms of the number of news segments per day, duration of news per day, 2) is more likely to focus on the shooter rather than the shooting itself.

1.2.2.1 The Intensity of News Coverage

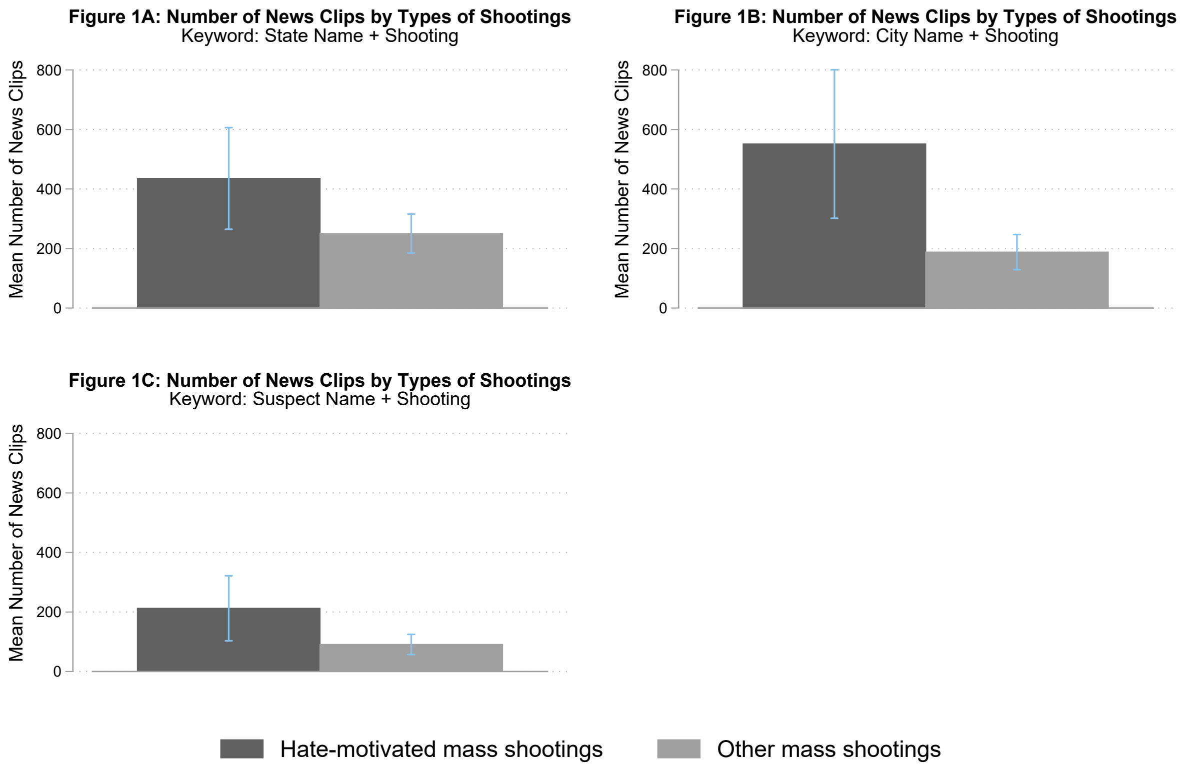
To compare the level of media coverage of hate-motivated mass shootings versus other mass shootings, I use two data sources. First, I collect data from The Internet Archive, a digital library that archives clips of U.S. television news broadcasts since 2009. For each mass shooting in my dataset from 2009 to March 2021, I search for news by captions using three different keywords: <State Name Shooting>, <City Name> Shooting, <Suspect Name> Shooting.[11] I restrict the search to include only news clips that are published within 7 days since the shooting happened. My outcome of interest is the number of news clips returned from searching.

Figure 1.1 displays the total number of news clips in the 7 days following the shooting. Across all three specifications, hate-motivated mass shootings are associated with significantly higher number of news clips. This suggests that on average, hate-motivated mass shootings receive more media coverage compared with other mass shootings.

Next, I use an alternative data source to validate the pattern in Figure 1.1. I collect data from

[11] I exclude mass shootings with no identifiable suspect names. My final sample size is 124 mass shootings.

Figure 1.1: Number of News Clips



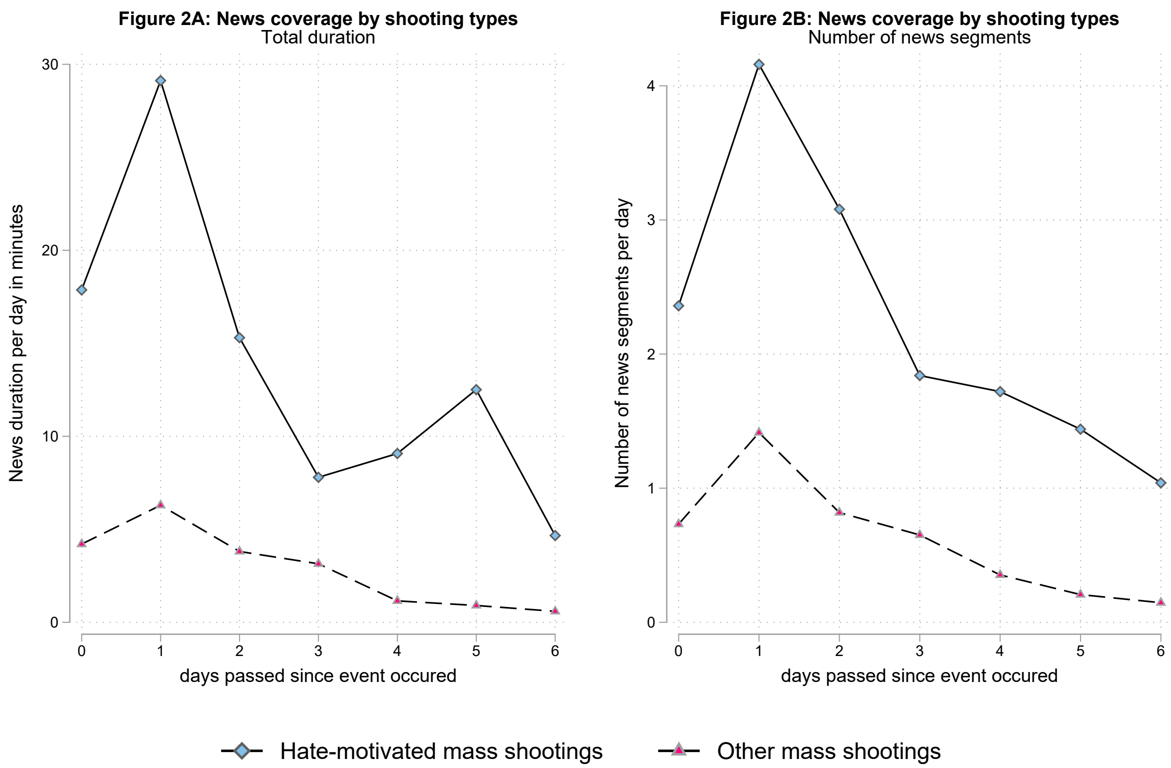
Notes: This graph plots the number of returned news clips when searching using the Internet Archive. The dark gray bar plots the average number of news clips for hate-motivated mass shootings, in the 7 days following the shooting. The light gray bar plots the average number of news clips for other mass shootings. Error bars reflect 95% confidence intervals. Only news that are published within 7 days since the shooting happened are included.

the Vanderbilt Television News Archive (VTNA) (Eisensee & Strömberg, 2007; Durante & Zhuravskaya, 2018; Jetter & Walker, 2022). VTNA's core collection consists of regularly scheduled evening newscasts from ABC, CBS, NBC, CNN. For each mass shooting in my dataset, I search for its news coverage within 7 days the shooting took place. To maximize accuracy, I use the suspect's name as a keyword.[12] This returned 1289 news clips for 223 mass shooting events.[13] I then construct a panel data of media coverage: I aggregate the total minutes of news coverage a shooting event received for each day within 7 days of the event. I then investigate the difference in

[12]I excluded mass shootings with no identifiable suspect names. My final sample size is 223 mass shootings.
[13]Note that the number of clips is much smaller than Internet Archive because VTNA only contains national newscasts.

the amount of coverage between hate-motivated mass shootings and other mass shootings.

Figure 1.2: Total Minutes on Evening TV News Broadcasts



Notes: This graph displays the intensity of media coverage of mass shootings using data from VTNA. Panel A plots the total duration of news in minutes. Panel B plots the number of news segments per day. The X-axis represents the number of days passed since the shooting occurred. The diamond solid line plots the average media coverage for hate-motivated mass shootings while the triangle dashed line does the same for other mass shootings.

Figure 1.2A shows that on average, hate-motivated mass shootings receive more news coverage per day, measured by the total duration of news in minutes. Figure 1.2B shows that on average, hate-motivated mass shootings are associated with higher number of news segments per day.

I formalize the graphical evidence in Figure 1.2 using regressions. Appendix Table A.3 shows the estimated coefficients under different specifications. My dependent variable is the total amount of news coverage per day measured by minutes. Estimates from regressions confirm the graphical evidence presented earlier. The coefficient on Hate is positive and statistically significant even after controlling for casualty level and a rich set of fixed effects. Estimates from column 4 show that when there is no death, hate-motivated mass shootings on average receive 6.5 more minutes of news coverage per day. For every additional dead victim, hate-motivated mass shootings receive 0.72 more minutes of news coverage while non-hate-motivated mass shootings receive only 0.21 more minutes.

1.2.2.2 The Content of News Coverage

In this subsection, I present evidence that media coverage of hate-motivated mass shootings differs in content compared with other mass shootings: news stories on hate-motivated mass shootings tends to focus more on the shooter.

I collect data from Google News, one of the world's largest news aggregators. For each mass shooting in my dataset that happened after 2006 (Google News was launched in 2006), I scrape down 50 news articles on Google News sorted by relevance.[14] I restrict my sample to articles that are published within 14 days after each shooting. I widened the scope from 7 days in prior analysis to 14 days because for a given event, there tends to be higher number of written news articles than the number of Television appearances. I use Octoparse to scrape down URLs and Python library Newspaper3K to extract data from each URL. This gives me 4396 valid news articles from various News agencies.[15] The top 3 sources are CNN, the New York Times, and the Washington Post. For each news article, I use NLP (Natural Language Processing) to obtain a set of keywords summarizing the article. I identify if a news article is about the shooter by examining whether the keywords of the article contains "suspect/shooter/gunman/perpetrator/killer/motive." As a robustness check, I do the same exercise using the title of the news article instead.

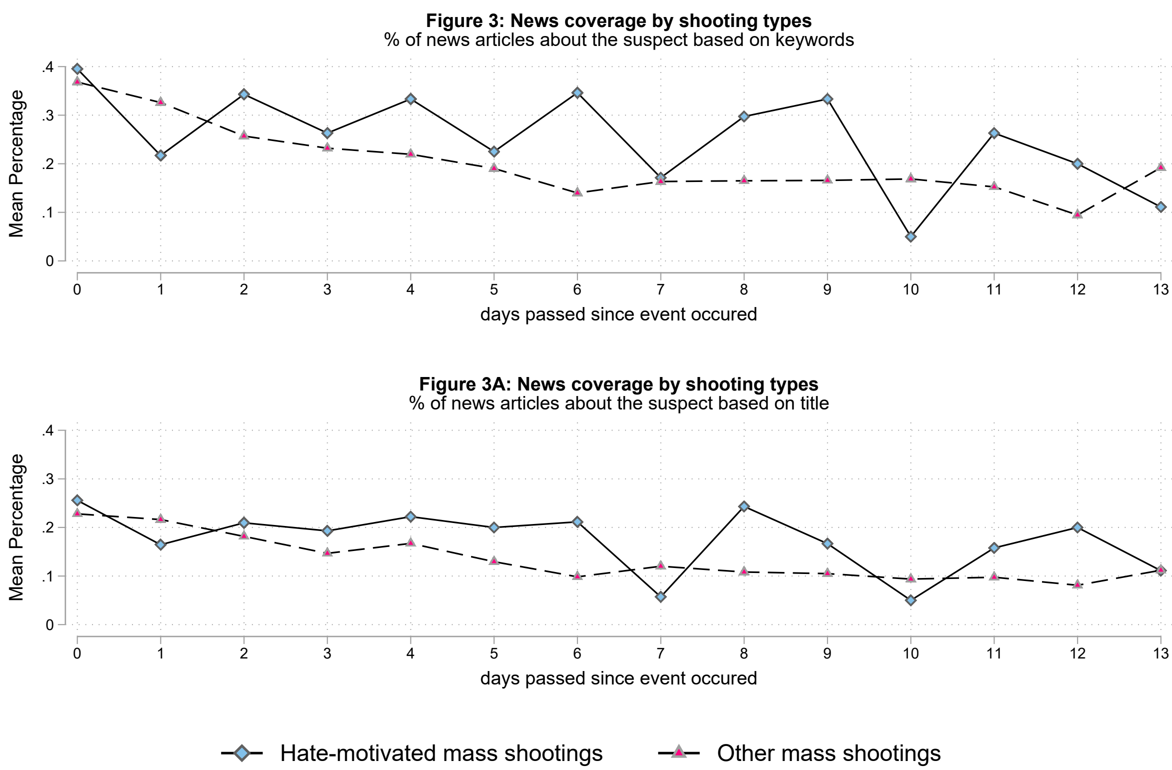
Figure 1.3 plots the fraction of news articles about the suspect each day based on the article keywords and title respectively. The two approaches produce similar patterns. News articles about a hate-motivated mass shooting are more likely to involve stories about the shooter. Appendix Ta-

[14] The number 50 is chosen based on an eyeball test of relevancy for returned results (the search might return irrelevant news articles) and computational constraints.

[15] I scraped down 5909 news articles. However, some are excluded from analysis due to missing information.

ble A.4 shows the difference in regression format. The dependent variable is an indicator variable that equals 1 if the news article focuses on the shooter. Estimates show that the difference is robust to controls and fixed effects. Column 4 shows that news articles on hate-motivated mass shootings are 5 percentage points more likely to focus on the shooter, which is a sizeable gap (21.28% increase) considering the proportion of news articles that involves shooters in non-hate-motivated mass shootings is 23.5 percent.

Figure 1.3: Percentage of News Articles about the Shooter



Notes: This graph displays the fraction of news articles that focuses on the shooter. Panel A uses the titles of news articles for identification. Panel B uses keywords. The X-axis represents the number of days passed since the shooting occurred. The diamond solid line plots the fraction for hate-motivated mass shootings while the triangle dashed line plots the fraction for other mass shootings.

The next question in line is, what exactly do the news stories talk about when they feature the shooter? I provide evidence in Appendix Table A.5 that more than 9% of all the new articles are related to the shooter's ideology/motive, when the shooting is hate-motivated. In comparison, when the shooting is not motivated by hate, the fraction drops to less than 2%. Consider the recent Buffalo shooting on May 2022. After the shooting occurred, media immediately started reporting stories about the shooter and his white supremacy ideology. This is disturbing since this would translate to an increase in exposure to the shooter's hateful ideology. Based on literature in criminology and psychology (see Lankford & Madfis, 2018a for a review), this could lead to behavioral contagion and other negative effects.

1.2.3 Reactions to Mass Shootings in the US

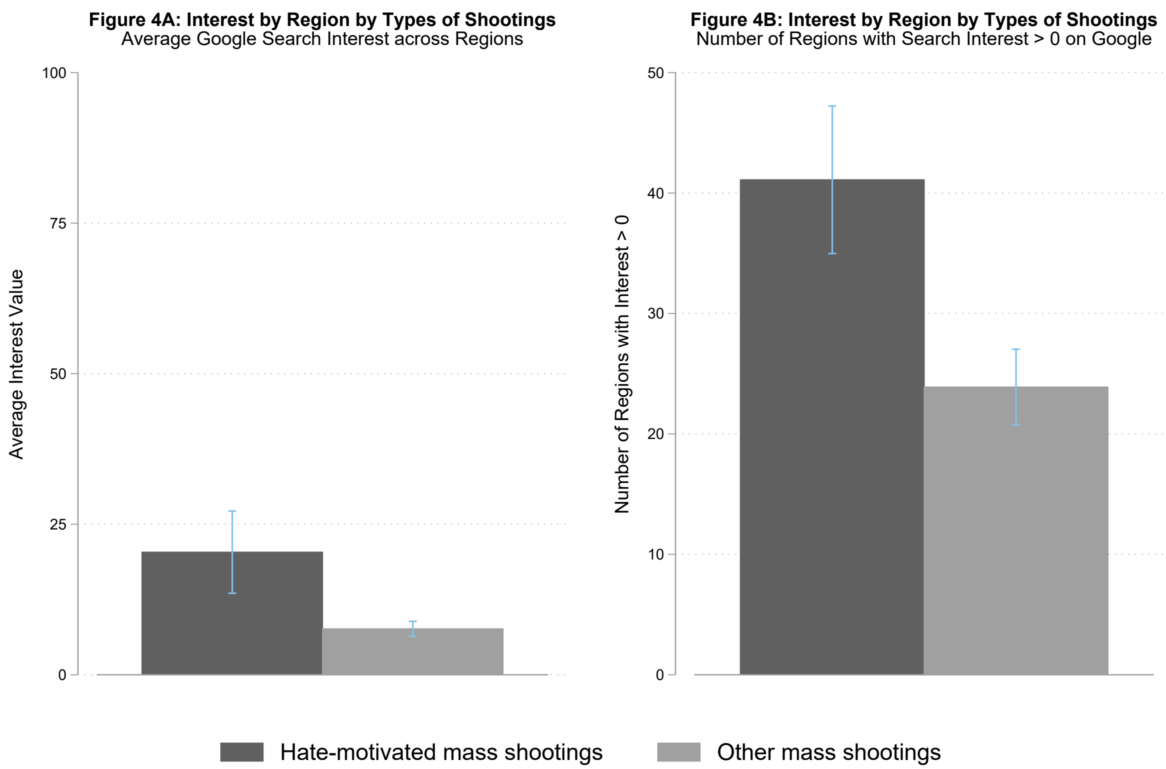
In this last subsection, I showed that from the media's side, hate-motivated mass shootings receive substantially higher news coverage compared to non-hate-motivated mass shootings. I now check whether similar differences exist from the viewer's side, i.e., whether people are more interested in hate-motivated mass shootings, and more specifically, whether people are more likely to look for information about the shooter. I collect data from Google Trends. Google Trends provides access to search requests made to Google. The data is aggregated and normalized.[16] For each mass shooting in my dataset that happened after 2006 (Google Trends was launched in 2006), I download data from Google Trends using Python library pytrends. I restrict the time frame of data to reflect searching behaviors in the United States within 14 days of each shooting.

My first outcome of interest is the state-level search interest in each mass shooting. I plot the average number of regions with non-zero search interest and the average search interest across all the regions in Figure 1.4. The graph shows that people are much more interested in hate-motivated mass shootings. Regression estimates in Appendix Table A.6 confirm this results. I use similar specifications as before. The dependent variable is the search interest of a mass shooting in a subregion. Column 4 shows that after controlling for the number of victims and fixed effects, the search interest value for hate-motivated mass shootings are on average 7.33 points higher compared to non-hate-motivated mass shootings. For every additional dead victim, the search interest for

[16]Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics.

hate-motivated mass shootings increases by 0.74 while the search interest for non-hate-motivated mass shooting increases by only 0.4.

Figure 1.4: Interest in Mass Shootings Measured by Online Searching Behavior



Notes: Google divides the United States into 51 subregions based on geography. For each subregion, Google calculates a search interest value, ranging from 0 to 100. A value of 100 would indicate in that subregion, the total search volume related to a given shooting divided by the total search volume is the highest in the United States. A value of 50 would indicate a subregion where searches about a given shooting are half as popular as the location that has a value of 100. Panel A shows the avearge search interest value across all subregions. Panel B shows the average number of subregions with a search interest value that is greater than 0. The data is restricted to reflect searching behaviors in the United States within 14 days since each shooting happened. The dark gray bar represents hate-motivated mass shootings while the light gray bar represents non-hate-motivated mass shootings. Error bars reflect 95% confidence intervals.

My second outcome of interest is people's subject of interest when searching information for mass shootings. Recall that in section 1.2.2, I showed that news coverage on hate-motivated mass shootings are more likely to focus on the shooter. I now examine whether people shower higher search interest in the shooter when the shooting is hate-motivated. For each mass shooting event, I retrieve a list of related topics. These words and phrases are the most common topics that the users who searched for the mass shooting also searched for during the same search session.[17] Figure 1.5 displays the fraction of mass shootings that has "Suspect" on its list of related topics.[18] Whenever a hate-motivated mass shooting happens, about 60 percent of the time, "Suspect" is among the most searched topics on Google. In contrast, this ratio drops to less than 30 percent when the shooting is not hate-motivated. The regression analysis in Appendix Table A.7 confirms this difference. However, the estimates lose statistical significance after adding control variables and fixed effects. In Appendix Table A.8, I provide evidence that people often search for keywords related to the shooter's motive. This search preference matches the media's tendency to report more about the shooter and the shooter's motive in hate-motivated mass shootings.

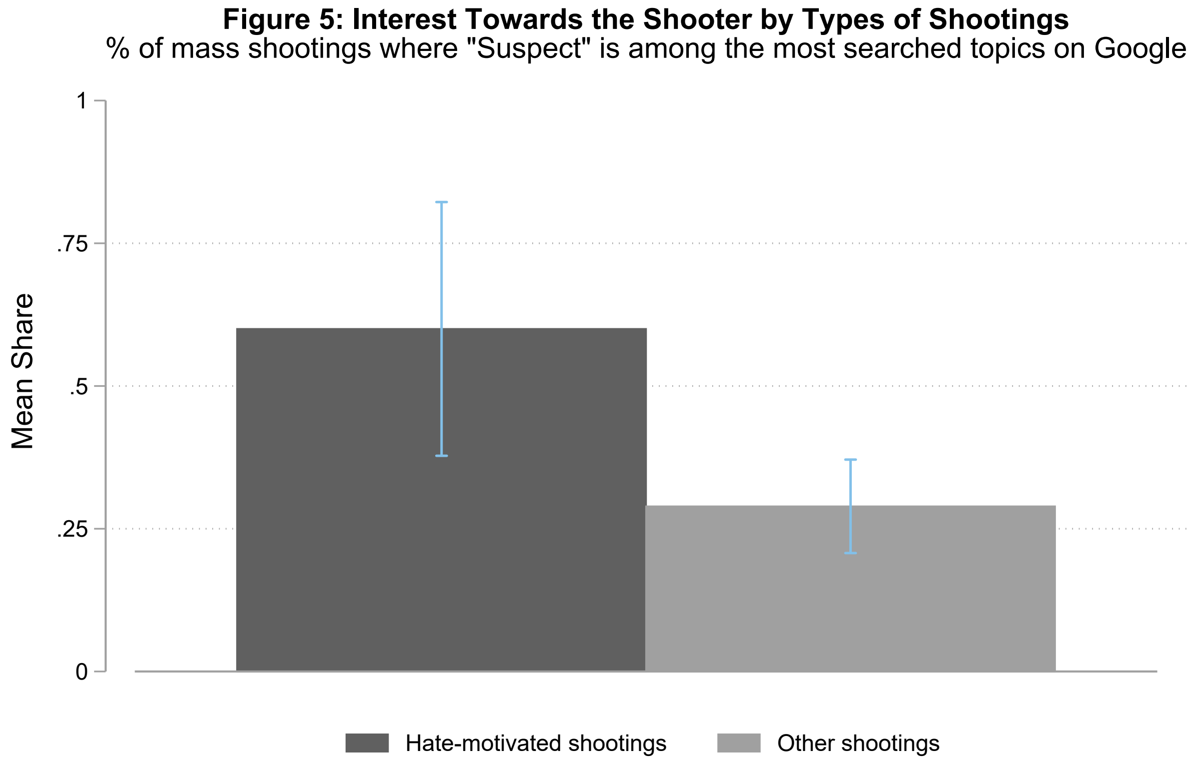
1.2.4 Hate Crime before and after Hate-motivated Mass Shootings

In Section 1.2.3, I showed that people show substantially more interest in the shooters of hate-motivated mass shootings, as reflected in online searching behaviors. This pattern matches the media's reporting pattern shown in Section 1.2.2. Assuming that this prolonged exposure to the shooter leads to behavioral contagion, e.g., people copying the shooter, I should expect to see an increase in hatred toward the victimized group targeted by the shooter. In this section, I investigate this possibility by estimating an event study model that compares how the number of hate crimes targeting one group changes over time from 10 days before the occurrence of a hate-motivated mass shooting targeting that same group to 10 days after the shooting. I use the hate crime data from the FBI. The data is incident level and includes every hate crime incident from 1991 to 2019. During the same period, there are 21 hate-motivated mass shooting. I aggregate the data by date (10592 days) and calculate the number of hate crime incident per day nationwide. I then estimate

[17] Google does not explicitly define what a search session is. Generally, a search session consists of all the search requests from a user within a given timeframe. A session lasts until there is inactivity. A common value for the inactivity threshold is 30 minutes and is sometimes described as the industry standard.

[18] People who searched for hate-motivated mass shootings are also more likely to search for topics that are related to the victims of the shooting. In particular, people search for the group that the victims belong to, i.e., race, religion. For example, those who searched for the Atlanta shooting in 2021 also searched for "Asian people," "Asia," "Asian Americans," those who searched for the Poway shooting in 2019 also searched for "Synagogue," "Chabad," "Rabbi."

Figure 1.5: Interest in the Shooter Measured by Online Searching Behavior



Notes: This graph shows the fraction of mass shootings that has "Suspect" on its list of related topics. For each keyword, Google returns at most 25 topics sorted by popularity. A topic includes all search terms related to it. Thus, for each mass shooting, the list of related topics contains the most common terms that users who searched for the mass shooting also searched for during the same search session. The data is restricted to reflect searching behaviors in the United States within 14 days since the shooting happened. The dark gray bar represents hate-motivated mass shootings while the light gray bar represents non-hate-motivated mass shootings. Error bars reflect 95% confidence intervals.

the following specification via ordinary least squares:

$$y_{Gt} = \sum_{t=-5}^{t=5} \beta_t * HateShooting_t + \gamma_t + \epsilon_{Gt}$$

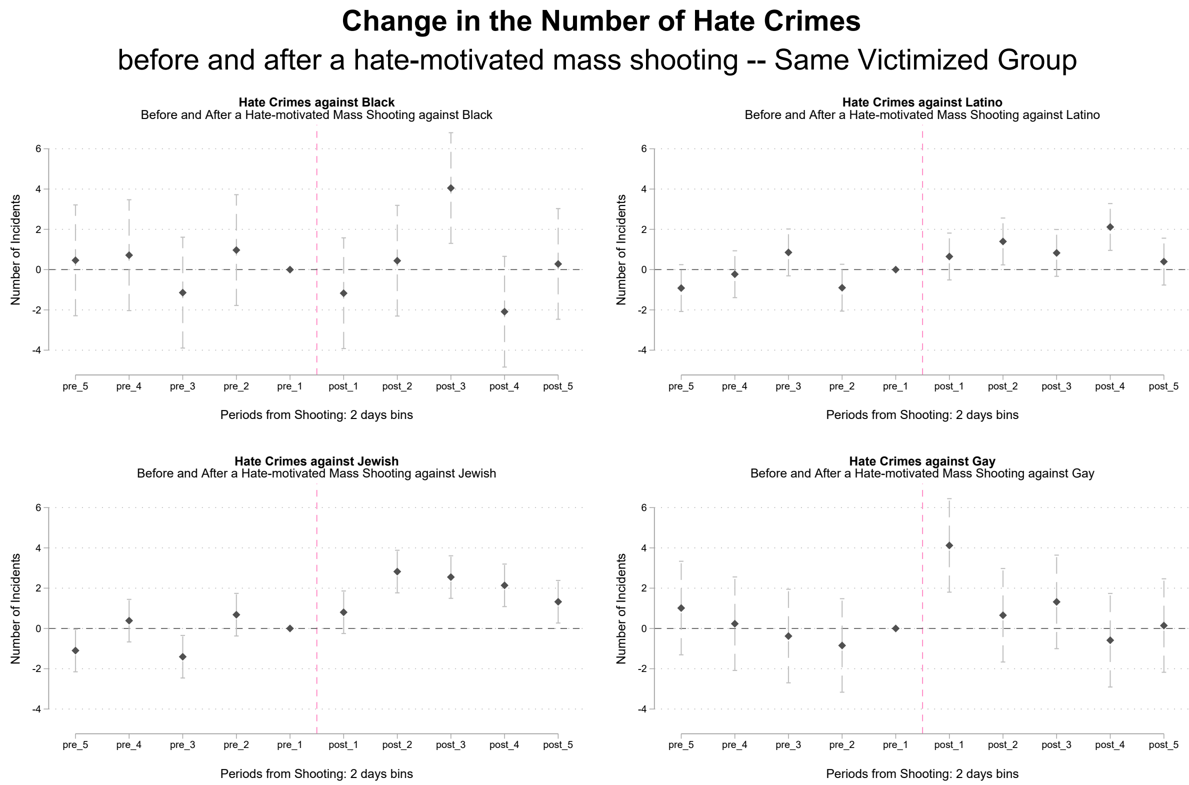
The outcome variable y_{Gt} is the number of hate crimes against group G on date t in the United States. $HateShooting_t$ are indicator variables for being within 10 days from a hate-motivated mass shooting. Each indicator variable represent a two-day interval, $t = -1$ is the omitted group. For example, if a hate-motivated mass shooting happened on date t , β_1 will measure the changes in the number of hate crimes on date $t+1$ and $t+2$. The set of coefficients β_k reflects the degree to which hate crime against group G changed before and after a hate-motivated mass shooting against group G. γ_t includes year fixed effect, month fixed effect, and day of the week fixed effect to control for changes over time in the rate of hate crimes. To examine possible heterogeneity by the types of mass shooting and hate crime, I investigate four different cases, the victimized group is African American, Latino, Jewish, and Gay respectively.

Figure 1.6 plots the resulting β_k estimates. The pink dotted line represents day 0 when a hate-motivated mass shooting happened. As the null estimates left of the link dotted line indicate, once all the fixed effects are controlled for, there is little difference in the pre-shooting trend in the number of hate crimes. Following a hate-motivated mass shooting, there's a large and statistically significant increase in the number of hate-crimes against the same victimized group targeted in the hate-motivated mass shooting. Consider the graph on the bottom left, the point estimate indicates that, relative to day 1 and day 2 before a hate-motivated mass shooting against Jewish took place, the number of hate crimes against Jewish increased by 3. This magnitude is large given that the daily average number of hate crime against Jewish is less than 3. The same pattern is also evident from the number of hate crimes against Black, Latino, and Gay, following a hate-motivated mass shooting against Black, Latino, and Gay respectively.

As a robustness check, I investigate whether the same pattern exists for the number of hate crimes against different victimized groups, i.e., following a hate-motivated mass shooting against African Americans, will there be an increase in the number of hate crimes against non-African Americans? To do so, I estimate a similar specification, except that the outcome variable is now the number of hate-crimes against non-G on a particular day. Figure 1.7 plots the resulting β_k estimates. None of the estimate coefficients for post periods is significant, suggesting that the number of hate crimes against different victimized groups does not change before and after a hate-motivated mass shooting.

While my findings strongly suggest that following a hate-motivated mass shooting, there is an increase in hatred toward the group that was targeted in the shooting, there are limitations to

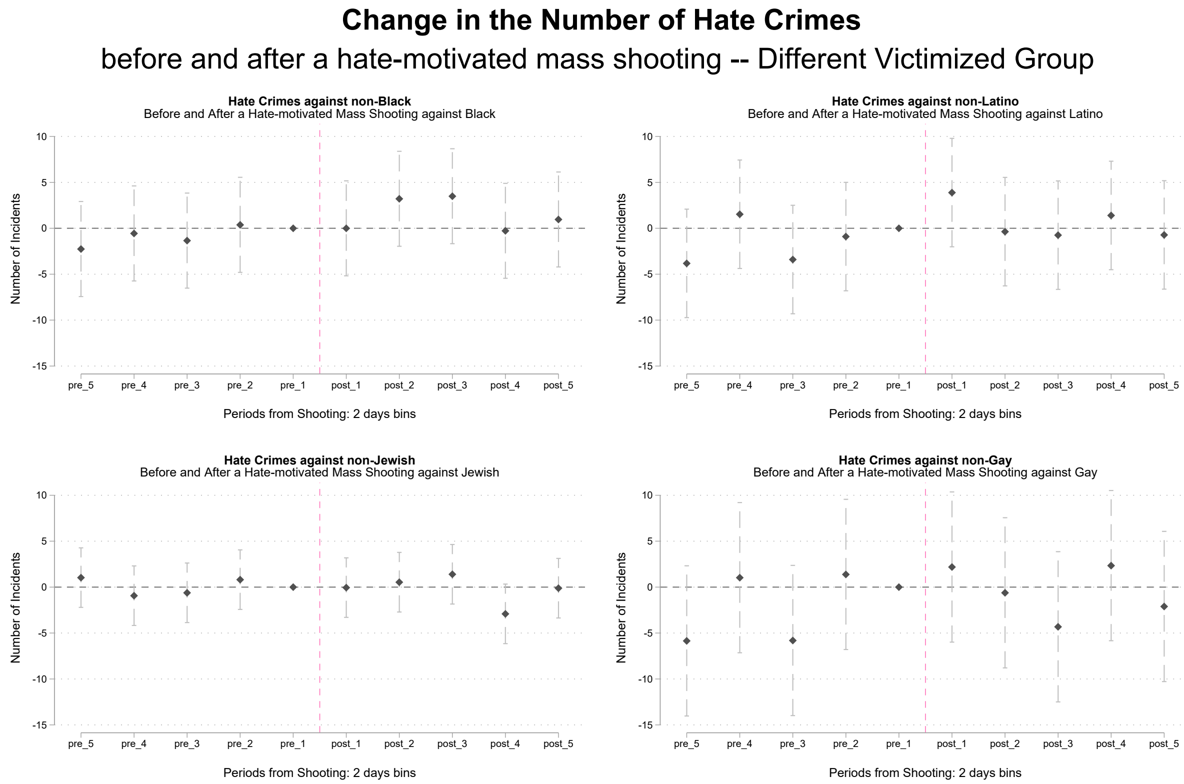
Figure 1.6: Change in the Number of Hate Crimes targeting the Same Population



Notes: This graph plots the estimated β_k coefficients as explained in Section 1.2.4. The dependent variable in Panel A, B, C, D is the daily number of hate crimes nation wide against African Americans, Latino, Jewish, and Gay respectively. The data is incident-level hate crime data from the FBI. The pink dotted line represents day 0 when a hate-motivated mass shooting happened. Each point on the X-axis represents a 2-day interval. For example, $post_1$ covers day 1 and day 2. pre_1 is the omitted group. I additionally control for year fixed effect, month fixed effect, and day of the week fixed effect, and the intensity of news coverage on each hate-motivated mass shooting 7 days following the shooting. Error bars reflect 95% confidence intervals.

my analyses. To begin with, with existing data, I can not prove that the increase in hatred is caused by news coverage of hate-motivated mass shootings. It is entirely possible that media is not responsible. For example, while people display higher interest in the shooters of hate-motivated mass shooting as shown in Section 1.2.3, this could simply reflect people's natural interest, which may have nothing to do with media coverage. The increase in hate crimes could be attributed to other reasons as well, such as victims' increased willingness to report, or the police's increased effort. In addition, FBI's hate crime data could be biased in two ways. First, as with most self-

Figure 1.7: Change in the Number of Hate Crimes targeting Different Populations



Notes: The estimates presented in this graph serve as a robustness check for the results shown in Figure 1.6. The dependent variable in Panel A, B, C, D is the daily number of hate crimes nation wide excluding the ones targeting African Americans, Latino, Jewish, and Gay respectively. The data is incident-level hate crime data from the FBI. The pink dotted line represents day 0 when a hate-motivated mass shooting happened. Each point on the X-axis represents a 2-day interval. For example, $post_1$ covers day 1 and day 2. pre_1 is the omitted group. I additionally control for year fixed effect, month fixed effect, and day of the week fixed effect, and the intensity of news coverage on each hate-motivated mass shooting 7 days following the shooting. Error bars reflect 95% confidence intervals.

reported crime data, it suffers from under reporting (Pezzella et al., 2019). Second, It is sometimes difficult to determine whether an incident is a hate crime. The labeling process could be subjective to biases from law enforcement agencies.[19] Therefore, observational data is not sufficient to show media coverage of hate-motivated mass shootings causally generate more hatred.

[19]The following is taken from the FBI's webpage. "*Only when a law enforcement investigation reveals sufficient evidence to lead a reasonable and prudent person to conclude that the offender's actions were motivated, in whole or in part, by his or her bias, should an agency report an incident as a hate crime.*"

1.2.5 Summary

In sum, using observational data from multiple sources, I found evidence that news coverage of hate-motivated mass shootings and non-hate-motivated mass shootings differ in two aspects. First, hate-motivated mass shootings receive more news coverage, as measured by the number of news clips, and the total length of coverage. Second, news coverage of hate-motivated mass shootings tends to focus more on the shooter such as the shooter's ideology. Next, I showed that people's reactions to hate-motivated mass shootings and non-hate-motivated mass shootings differ in similar manners. Namely, based on online searching data, I find that people show a higher interest in hate-motivated mass shootings, and this is possibly driven by the higher interest in the shooter. Finally, if the increase in exposure to the shooter results in an increase in hatred, then following a hate-motivated mass shooting, there should be an increase in the number of hate crimes against the same group that was targeted in the shooting. Using the FBI's hate crime data and an event study framework, I provide evidence that there is indeed a rise in hatred post-shootings.

Based on this evidence and the existing literature, I hypothesize that the news coverage of hate-motivated mass shootings have unintended consequences, namely, it generates more hatred toward the victimized group. However, to provide causal evidence, one would have to overcome identification challenges. The ideal natural experiment would require exogenous variations in the way media covers mass shootings. This is extremely hard to achieve with observational data, making the implementation of an experiment especially desirable. In the next section, I describe the design and implementation of an online survey experiment, and lay out my hypotheses.

1.3 Experimental Design

Based on my findings from observational data and the existing literature, I design and conduct an online information provision experiment to causally examine the impact of news coverage of hate-motivated mass shootings. My main treatment conditions simulate media's tendency to focus on the shooter when covering hate-motivated mass shootings. In particular, I examine two aspects, 1) media's focus on the shooter's ideology, and 2) media's focus on the shooter's identity and background. This section introduces experimental design, treatment manipulations, hypotheses, outcome measures, and estimation strategy.

1.3.1 Outline

In order to causally study the impact of news coverage of hate-motivated mass shootings, I design and conduct an online information provision experiment. The structure of my experiment largely follows the growing literature on information provision experiments (for a review, see Haaland et al., 2020). This methodology has been applied to answer policy-relevant questions in a variety of fields including sensitive topics such as attitudes toward immigration, labor market discrimination, and xenophobia (Alesina et al., 2018; Bursztyn, Egorov, & Fiorin, 2020; Grigorieff et al., 2020; Haaland & Roth, 2020). This methodology is ideal for my research for three reasons. First, I can exogenously vary the content of the news story on a mass shooting, generating the ideal variation I need to test my hypotheses. Second, I can vary one feature of the information set at a time to cleanly identify what specific content of the news story is affecting subjects' attitudes. Third, experiments conducted online provides a higher degree of anonymity compared with experiments conducted in the laboratory. Thus, participants in online experiments should feel more comfortable expressing their true views.[20]

In the experiment, participants are asked to read a piece of news story about a hate-motivated mass shooting against Hispanic immigrants. I implement different treatments where the content of the story is experimentally manipulated to either emphasize or do not emphasize the hateful nature of the shooting, the hateful ideology of the shooter, and the identity and background of the shooter. I subsequently measure participants' interest in the shooter's ideology and background, their attitudes toward the shooter (including admiration and justification for the shooter's action), their support for the hateful ideology of the shooter, and their interest in accessing information regarding a white supremacy hate group.

The news story in my experiment focuses on the 2019 El Paso shooting. This shooting has been

[20]I formally address issues related to social desirability bias in Section 1.4.4.

described as the deadliest anti-Latino and anti-immigrants attack in recent U.S. history.[21] Twenty three people were killed in the shooting, including 8 Mexicans. Due to the high level of casualty, it received extensive media coverage in the United States. I choose this shooting for three reasons. First, the hateful motives behind the shooting is evident. The shooter Patrick Crusius explicitly admitted in his manifesto, that the shooting was motivated by anti-immigrant and anti-Hispanic ideology. Thus, this shooting clearly falls under the hate-motivated mass shooting category defined in this paper. Second, while there are other well-defined hate-motivated mass shootings targeting different populations such as people of certain race or sexual orientation, people might feel reluctant to express their true attitudes toward such populations in the modern society, given the potential of backlashes.[22] In comparison, discussion about immigrants and immigration are less frowned upon compared to other sensitive issues, since policies regarding immigration and anti-immigration are an essential part of the political agendas for both Democrats and Republicans. For example, recent survey evidence shows that college students think it is easier to have an open and honest conversation about immigration than racial inequality and gender-related topics.[23] Moreover, there exists a variety of both anti- and pro-immigration organizations actively attempting to influence legislation.[24] With the rise of Donlad Trump, the anti-immigrant sentiment has been increasingly mainstreamed over the last few years.[25] Third, since immigration is an issue that is explicitly polarized based on political leaning with the Republican party known to take a stricter stance, I can use subjects' political affiliation as a proxy to measure ex-ante attitudes toward the victimized group in the shooting, i.e., immigrants.

Figure 1.8 shows the outline of the experiment. Upon consent, subjects are first asked a series of demographic questions. Next, subjects are randomly assigned to one of the four information

[21] For example, see this article published by the New York Times.

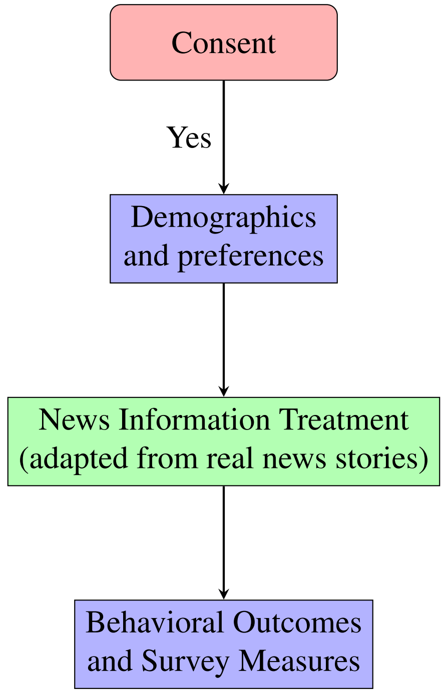
[22] For example, the National Association of Scholars complied a list of individuals who lost their job for offensive remarks. While the majority of cases are related to racism and gender, only less than 2% of the cases are related to immigrants and immigration.

[23] See the report on college free speech: Link. On a related note, Ekins (2017) finds that among a list of 13 controversial topics, the majority of Americans believe speakers should be allowed to talk about deporting illegal immigrants at college campus.

[24] For example, see this list from Columbia University.

[25] For example, see this report from ADL.

Figure 1.8: Outline of Experiment



treatments where they are asked to read a piece of news story about a real past mass shooting. Finally, I measure my outcomes of interest, including subjects' interest in the shooter, subjects' attitudes toward the shooter, subjects' attitudes toward the shooter's anti-immigrant ideology, and subjects' interest in information about a white supremacy hate group.

I take several measures to improve data quality. First, before subjects are presented with the news story about the mass shooting, they are told that they will be asked to read a real piece of news story randomly selected from three categories: business, crime, and sports.[26] This design choice intends to mask the true objective of the survey, thus reducing participants' beliefs that the study aimed to measure their bias against the victimized group. Second, I offer rewards to ensure that subjects read the news story carefully. Before subjects see the news article, they are told that they have to answer two comprehension questions about the story, and they would be compensated $0.2 for each correct answer. Third, in order to measure if subjects are paying attention, I include two attention checks in the survey following Oppenheimer et al. (2009), one pre-treatment, one

[26]To maximize statistical power, 99.9% of the subjects saw the crime version of the survey.

post-treatment. Subjects are aware that their submission might get returned if they fail attention checks.[27] Finally, it is possible that subjects have seen news stories about the El Paso shooting before. In order to reduce recall bias, dates and locations are omitted from all treatments, with the exception of Treatment 4 mentioning the shooter's name.[28]

1.3.2 Treatments

I implement 4 different treatment conditions, where I experimentally manipulate the content of the news story in a way that progressively increases information about victims, motives and the shooter's background. To mimic real-life news as closely as possible, the wordings in all treatment conditions are adapted from real stories posted by major US media outlets.[29] Detailed survey script including the news story in each treatment condition are included in Appendix A.2.

1.3.2.1 Treatment 1: No Hate

The No Hate Treatment represents the most streamlined news coverage of mass shootings, featuring a basic description of the event, including when and where the shooting happened, the number of victims, and excerpts from interviews with witnesses. This news coverage does not mention the hate motive behind the shooting and does not provide information about the victimized group. Instead, the shooter's motive is described as economic reasons according to a manifesto shared online.[30]

1.3.2.2 Treatment 2: Hate

The Hate Treatment is identical to Treatment 1, with the exception that the news story now mentions the hate-motivated nature of the shooting, i.e., victims include at least 8 Mexicans, the attack was targeted at the local Hispanic community, authorities are investigating possible hate crime charges. This treatment represents the minimalistic news coverage of hate crimes. Although

[27]For example, see Prolific's Attention and Comprehension Check Policy. An example of attention checks is provided in Appendix A.2.4.

[28]In the end of the survey, I asked subjects: "Before today, have you heard of the event described in the news story that you read?" 22.6 percent chose yes. I address this issue in detail in Section 1.4.4.3.

[29]Sources: NPR, The New York Times, NBC, BBC, The Washington Post

[30]This is partially true according to the shooter's manifesto.

the news story discloses the possibility of a hate crime, no further details are given regarding why the shooter targeted Hispanic immigrants. Recall that in Section 1.2.3, I showed that people are more interested when the mass shooting is hate-motivated. This could simply be a natural preference regardless of how media reports the story. The comparison of the No Hate Treatment and the Hate Treatment allows me to examine whether people naturally react to hate-motivated mass shootings differently, regardless of the content and intensity of media coverage.

1.3.2.3 Treatment 3: Hate & Ideology

The Hate Ideology Treatment builds on the Hate Treatment by adding details on the shooter's ideology by providing excerpts from the shooter's manifesto:[31]

> The document claims that the attack was targeted at the local Hispanic community. It stated that Latin America immigrants represented a "Hispanic invasion." It warned that white people were being replaced by foreigners.
>
> The manifesto described an imminent attack and railed against immigrants, saying, "if we can get rid of enough people, then our way of life can be more sustainable." It also detailed a plan to separate America into territories by race to save this country.
>
> The author hoped his/her attack and words would inspire additional like-minded attacks and lead to a wider racial violence in pursuit of a white ethnostate.

In Section 1.2.2.2, I provide evidence that media coverage on hate-motivated mass shootings tends to focus on the shooter and the shooter's ideology. In Section 1.2.3, I provide evidence that people show high interest in information related to the shooter and the shooter's ideology. The matching preferences suggest whenever a hate-motivated mass shooting happens, there will be a high exposure to the shooter's hateful beliefs and ideas. The comparison of the Hate Ideology Treatment and the Hate Treatment allows me to isolate the impact of media's emphasis on the shooter's hateful ideology on the audience's attitudes toward the shooter and the victimized group.

[31]The wordings are taken directly from the shooter's manifesto, and from an article published on the New York Times.

1.3.2.4 Treatment 4: Hate & Background

The Hate Background Treatment builds on the Hate Treatment by adding details on the shooter's identity and background, including his name, photo, and words from former classmates who saw the shooter depressed and bullied in high school:[32]

> Police officers were interviewing the suspect, Patrick Crusius, a 21-year-old white man from Allen, Tex.
>
> Investigators are looking into whether Crusius might have been radicalized online. But friends and former teachers and classmates say he might have been hardened, too, by the tensions in his changing community in real life.
>
> Allison Pettitt, a classmate, said she saw Crusius pushed around in the hallways and "cussed out by some of the Spanish-speaking kids." She said that bullying was common at the school and that teachers often ignored it. "He started getting more depressed closer to the end of junior year," Pettitt said. "He started wearing a trench coat to school and becoming more antisocial and withdrawn." Lesley Range-Stanton, a spokeswoman for Plano's school district, declined to comment about whether Crusius was bullied, citing student privacy.

The Hate Background Treatment investigates how a news story with emphasis on the shooter's identity and background story affects viewer's attitudes. While the attention given to the shooter from the media combined with the high interest from the public will likely lead to an increase of exposure to the shooter's ideology, it can also lead to an increase of exposure to other information related to the shooter. For example, people will likely encounter information about the identity and background of the shooter, including biography, life story, quotes from friends and neighbors' depiction. Although media often portrays the shooter from various angles, I choose a story that hints at the shooter's poor mental health condition because of its frequent occurrence on the news. The idea that there exists a link between mental illness and mass shootings is indeed often discussed

[32] The wordings are taken from an article published on The Washington Post.

in the media and in scholarly research. Despite increasing evidence that mental illness does not cause mass shootings, it is heavily debated. Depending on the criteria, the proportion of mass shootings associated with mental illness varies from 4.7% to 78% across studies (Parks et al., 2019). Consequently, news coverage of mass shootings and other violent events often features discussion about the offender's mental health (Swanson et al., 2015; McGinty et al., 2016; DeFoster & Swalve, 2018; Duxbury et al., 2018). A database complied by Mother Jones shows about 63.2% of mass shootings are linked to news stories about the shooter's mental health.[33] Therefore, the Hate Background Treatment closely resembles the way that news outlets cover mass shootings. In section 1.2.3, I provide evidence that people frequently search for information related to the shooter following a hate-motivated mass shooting. This suggests that people will very likely encounter news stories about the shooter's mental health sufferings either through direct news coverage or online searching. The comparison of the Hate Background Treatment and the Hate Treatment allows me to identify the impact of a stereotypical story about the shooter's childhood suffering may have on the audience attitudes toward both the shooter and the shooter's ideology.

1.3.3 Hypotheses

Building on existing literature and the patterns I found using observational data, I derive four hypotheses.

Hypothesis 1. *The Hate Ideology Treatment and the Hate Background treatment increases subjects' interest in the shooter.*

Media plays an important role in informing the public of criminal incidents. Recent survey shows Americans rate crime as one of the most important news topics for daily life (Center, 2019). In section 1.2.3, I showed that compared to non-hate-motivated mass shootings, people show significantly higher interest in the shooter in hate-motivated mass shootings. Moreover, this pattern

[33]This data can be accessed at `https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/`. For more references on media's portray of mass shooting and mental illness, see Klein, 2012; Saad, 2013; McGinty et al., 2014; Metzl & MacLeish, 2015; Fox & Fridel, 2016; Lankford & Cowan, 2020. Note that the FBI has been calling for the media to reduce coverage featuring the offender's life stories. However this effort seems futile. See here.

matches media's tendency to focus on the shooter when covering hate-motivated mass shootings. The Hate Ideology Treatment and the Hate Background Treatment mimic the media's emphasis on the shooter. Therefore, I hypothesize subjects in these two treatments will show higher interest in the shooting and the shooter.

Hypothesis 2. *The Hate Ideology Treatment increases subjects' support for the shooter's ideology.*

Existing studies in criminology and psychology argue that media coverage of mass shootings could generate negative consequences through behavioral contagion (see Lankford & Madfis, 2018a for a review), e.g., people might find inspiration from the shooter, endorse the shooter's ideology, and possibly also imitate the shooter's actions, leading to more crimes. In Section 1.2.2, I showed that whenever a hate-motivated mass shooting happens, news outlets tend to spend more time covering the shooter, including the shooter's ideology. In Section 1.2.3, I showed that people express more interest in the shooters of hate-motivated mass shootings. The combination of these two findings imply that, whenever a hate-motivated mass shooting occurs, the public is likely subject to an increased and prolonged exposure to the shooter's hateful ideology. While recent literature suggests that positive information about a minority group can improve people's attitude toward that minority group (Grigorieff et al., 2020; Haaland & Roth, 2020; Haaland & Roth, 2021; Settele, 2022; Song, 2022;), the effect of negative information such as racism is largely unknown. It is entirely possible that exposure to such information about a minority group can increase anti-minority beliefs and behaviors. If that is the case, knowing the shooter's anti-immigrant ideology could worsen people's attitudes toward immigrants. In section 1.2.4, using an event study framework, I showed that immediately following a hate-motivated mass shooting, there is an increase in hate crimes against the same victimized group. Building on these findings and the existing literature, I hypothesize that the Hate Ideology Treatment, which mimics media's emphasis on the shooter's ideology when covering hate-motivated mass shootings, will increase subjects' support for the shooter's ideology.

Hypothesis 3. *The Hate Background Treatment increases subjects' support for the shooter and the* *shooter's ideology.*

As discussed in the previous subsection, it is common to find news stories that link mass shooter's act to histories of abuse, being bullied, and difficult childhoods. Unsurprisingly, survey evidence shows the public blames mental health as the top reason for gun violence.[34] The Hate Background Treatment mimics media's emphasis on the struggling of the shooter. On the one hand, this may increase people's sympathy for the shooter. On the other hand, people might blame immigrants for the shooter's suffering and thereby increase their anti-immigrant sentiments. Taken altogether, I hypothesize that subjects in the Hate Background Treatment will show higher support for the shooter and the shooter's ideology.

Hypothesis 4. *The Hate Ideology Treatment and the Hate Background Treatment have differential impacts on subjects depending on their initial biases against immigrants.*

Conceptually, the Hate Ideology Treatment and the Hate Background Treatment can change people's attitude in three ways. First, emphasis on the shooter may serve as a persuasion device (DellaVigna & Gentzkow, 2010). The Hate Ideology Treatment introduces the shooter's hateful ideas to subjects and could persuade them to accept the ideas. Similarly, the Hate Background Treatment introduces the shooter's struggles and could persuade subjects to increase sympathy for the shooter. This effect should be more pronounced for subjects whose prior toward the victimized group is neutral, i.e., near the middle of the distribution. Second, emphasis on the shooter may serve as a coordination device (Arias, 2019). The Hate Ideology Treatment spreads the shooter's hateful ideas to all subjects and may thereby increase subjects' perceived popularity of the hateful ideas. In a similar fashion, the Hate Background Treatment could change subjects' perceived acceptability of supporting the shooter. Consequently, subjects with pre-existing biases may become more comfortable expressing support for the shooter and the hateful views given their updated beliefs. Taken together, I hypothesize that subjects with different initial bias against immigrants will respond differently to the Hate Ideology Treatment and the Hate Background Treatment.

[34] For example, click here and here to see the reports.

1.3.4 Outcome Measures

I have four primary outcomes: 1) interest in the shooter, 2) attitudes toward the shooter, 3) attitudes toward the shooter's anti-immigrant ideology, 4) interest in a white supremacy hate group.

Specifically, my experiment aims at answering three questions. First, does the way the media covers hate-motivated mass shootings cause an increase in people's interest in the shooting, the shooter and the shooter's ideology? Second, does the way the media covers hate-motivated mass shootings affect the audience's attitudes toward the shooter and/or the shooter's ideology? In particular, by emphasizing the hateful ideology or by disclosing the shooter's identity and background, could media coverage induce people to justify the crime and sympathize with the shooter, or to express support for the hate ideology, possibly leading to more hatred toward the victimized group? Third, does the effect of media coverage on hatred depend on people's pre-existing views regarding the victimized group?

1.3.4.1 Outcome 1: Interest in the Shooter's Ideology and Background

My first outcome attempts to address *Hypothesis 1*, i.e., whether media coverage of hate-motivated mass shootings causally increases the audience's interests in the shooter. After subjects read the news story, I ask them if they would like to receive full access to the shooter's manifesto and if they would like to receive more information on the shooter's identity and background. Both questions are binary, subjects are told that if they choose yes, they would receive access to the information at the end of the survey (Chopra et al., 2022).

There are several reasons that make this outcome a valid behavioral measure. First, in a real-world setting, it is typical that readers would see the title or a preview of the story first and then decide to read the story, or search for additional information, as documented by the google search data I analyzed in Section 1.2.3 of this paper. Second, since subjects are paid a pre-specified reward upon completion of the survey, by choosing to receive more information, subjects are committing to spending more time on the survey, therefore paying a cost to access the additional information on the shooter. Third, one could argue that subjects interested in the shooter's manifesto and background story could access the information provided at the end of the survey independently. They could search for such information by themselves after completing the survey, especially in the Hate Background Treatment where the identity of the shooter was disclosed. However, searching for information outside the survey would come at an additional cost (of time). Therefore, it is reasonable to believe that interested individuals would prefer to access such information while participating in the study.

1.3.4.2 Outcome 2: Attitudes toward the Shooter

To test whether the Hate Background Treatment increases subjects' support for the shooter, as specified in *Hypothesis 3*, I ask subjects three survey questions that measures their support and sympathy for the shooter: 1) whether they admire the shooter's courage, 2) whether they think the shooter's action can be justified, 3) what is the appropriate sentencing for the shooter. Admiration and justification are asked on a 5-point Likert Scale. Sentencing options range from 10 years or less imprisonment to the death penalty. The exact survey script is provided in Appendix A.2.

Mass shooters often regard previous shooters as role models or inspiration. Langman (2018) shows this connection is mainly formed through heroic idolization and shared sympathy. Thus, I design my attitudinal measures to capture these two aspects. In addition, having three measures allows me to cross check the identified treatment effects. One might be concerned that having a group of similar outcomes is susceptible to bias due to multiple hypothesis testing. To address this concern, I compute a standardized index of support, as further explained in Section 3.3.5.

1.3.4.3 Outcome 3: Attitudes toward the Shooter's Ideology

I test *Hypothesis 2* and *Hypothesis 3* by measuring subjects' attitude toward the shooter's ideology (Anti-Hispanic, Anti-immigrant). Ideally, my goal is to measure hatred towards the victimized group and willingness to commit crimes against them. However, this is challenging from both a design and ethical perspective. As a workaround, I construct a behavioral measure of subjects' support for anti-immigration. To do so, subjects are told that they will be given the opportunity to authorize a $1 donation to one randomly chosen organization. The organization will either be an anti-immigrant or a pro-immigrant organization.[35] Subjects see a brief description of only their randomly assigned organization before authorizing to donate. Subjects are also explicitly told that if they authorize the donation, the $1 will not be deducted from their payoff, they are simply authorizing it. This donation setup follows the methodology introduced by Bursztyn, Egorov, and Fiorin (2020) in the context of measuring anti-immigrant sentiment.[36] An alternative design would be to ask subjects to make donations out of their own money. However, given the sensitive nature of the causes supported by the organizations of interest, subjects might feel more reluctant to make the donation.

In the experiment, about 75% of the subjects are assigned to The Federation for American Immigration Reform (FAIR, an anti-immigrant organization), while about 25% of the subjects are assigned to American Immigration Council (AIC, a pro-immigrant organization). The primary outcome of interest is the willingness to authorize donations to FAIR. To make sure that donating to FAIR captures more of a resentment and hatred toward immigrants instead of reasonable support for stricter immigration policy, I made it clear in the description of FAIR, that this is an anti-immigrant organization with ties to white-supremacy groups and has made many racist comments. The pro-immigrant organization and the randomization of subjects to be shown one of the two organizations serve three purposes. First, it reduces experimenter demand effects by masking the true intent of the experiment. Subjects are told that one organization will be randomly selected and only see the chosen organization. Thus, it should be more difficult to relate the experiment to anti-immigrant sentiment. Second, the donation to the pro-immigrant organization serves as a robustness check. Assuming that the support for anti-immigration and pro-immigration are mutually exclusive, then the change in donation rate to the two organizations should move in opposite directions. Thus, if one treatment condition increases the donation rate to the anti-immigrant organization, it should decrease the donation rate to the pro-immigrant organization. Measuring

[35] In the experiment, subjects know that the organization will be randomly selected, but do not know it is restricted to two immigration organizations.

[36] Donation to organizations is widely used as an outcome measure in information provision experiment. For example, see Alesina et al., 2018; Bursztyn, Egorov, & Fiorin, 2020; Bursztyn, Haaland, et al., 2020; Grigorieff et al., 2020; Haaland & Roth, 2021.

subjects' donation to the pro-immigrant organization allows me to cross check the treatment effect on support for anti-immigrant. Finally, it is possible that the news coverage of hate-motivated mass-shootings has differential impact on attitudes towards the shooter and attitudes toward the victims. The emphasis on the hate ideology of the shooter may, for instance, increase support for the shooter from some individuals, and increase support for the victimized group from other individuals. My design allows me to test for this possibility.

1.3.4.4 Outcome 4: Interest in a White Supremacy Hate Group

To further test *Hypothesis 2* and *Hypothesis 3*, I construct a more direct measure of hatred by measuring subjects' interest in a white supremacy hate group. Subjects are given a brief description of a hate group named Stormfront, the oldest and one of the largest hate sites, then asked if they want to receive links to access its website. If subject expresses interest, they will receive links at the end of the survey. I then track whether subjects clicked on the provided links.[37]

This measure complements the donation outcome for several reasons. First, the interest in hate group is a closer proxy for hatred. Many shooters wrote and post manifestos on online hate groups before committing the shooting. Second, it is possible that subjects hate immigrants but do not donate to an anti-immigrant organization. This might be because they do not trust the organization, might be because they do not think a $1 donation will make a difference. In comparison, subject can access the hate group and "do things" by themselves. Third, it is possible that the treatment effects do not immediately translate into instant action. Many past mass shooters admitted that they are radicalized online.[38] The interest in hate group can capture this intermediate effect.

One might be concerned that subjects may simply request and click on the links out of curiosity rather than hatred. To minimize this concern, I make sure that subjects are aware that this hate group is founded by long-time white supremacist and is actively promoting white supremacy. If requests and clicks are driven by curiosity, then I should expect to see a large volume of requests

[37]The clicking is tracked using JavaScript embedded in the Qualtrics survey. Subjects are not aware that their clicking can be tracked.

[38]For example, the 2022 Buffalo supermarket shooter references a 4chan board devoted to guns, and says he was radicalised by the /pol/ or "politically incorrect" board. For further information see this article published by BBC.

and clicks. I address this issue again in Section 1.4.2.4.

1.3.4.5 Other Survey Measures to Detect Possible Mechanisms

My secondary analysis aims at identifying the mechanism behind the treatment effects. My treatments of interest are the Hate Ideology Treatment and the Hate Background Treatment, the news stories presented in both treatments provide extra information about the shooter. As discussed in *Hypothesis 4*, news coverage that focuses on the shooter can affect people's attitude through two channels. First, news stories that provide details on the shooter's anti-immigrant ideologies may persuade viewers into accepting the shooter's beliefs. Similarly, news stories that provide details on the shooter's struggles (traumatic childhood, bullied) may persuade viewers into sympathizing with the shooter. Second, news stories may change people's perception about the local popularity of anti-immigrant sentiment. Thus, people with an initial bias towards immigrant might find it more comfortable expressing support for the shooter and the shooter's ideology.

First, to examine whether news stories about the shooter can persuade people with no initial bias against immigrants to hate immigrants, I measure subjects' baseline attitudes before the information treatment. I elicit subjects' opinions on 6 political issues including abortion, same-sex marriage, gun control, minimum wage, build the wall, and citizenship for children of illegal immigrants.[39] Although my primary interest is in the two immigration-related questions, the four questions on other political issues aim to reduce experimenter demand effect. It also helps me evaluate the subject's political stance more generally. Detailed survey script or provided in Appendix A.2.4.1. If the persuasion mechanism is present, I should expect to see less treatment effects for subjects with strong initial bias since they are already persuaded.

Second, to examine whether news stories about the shooter can change subjects' perceived popularity of the shooter and the shooter's ideology, I elicit subjects' perception of social norms. After subjects complete the survey module on the attitudes toward the shooter described in Section 1.3.4.2, all subjects except for the first 200 are asked to guess the option that was chosen the most

[39]These topics are taken from popular political issues on iSideWith.com.

by the first 200 subjects. Subjects are paid \$0.2 for each correct guess.[40] Similarly, I elicit subjects' perception of the popularity of anti-immigrant sentiment. After subjects complete the donation question described in Section 1.3.4.3, they are asked to report what they think is the percentage of previous subjects who authorized the donation. Subjects are rewarded \$0.2 if the difference between their guess and the true answer is less than or equal to 2. If the treatment effects operate through the social norm channel, I should expect subjects' perception of social norms to move in the same direction as their actual behavior across different treatments.

Although the two mechanisms could coexist and it is difficult to isolate them, these survey questions allow me to generate insights on what may be driving the treatment effects.

1.3.5 Sample and Procedure

I recruit survey participants from two online recruiting platforms, Prolific and CloudResearch. Studies that compare data quality of behavioral research across online platforms consistently find that Prolific and CloudResearch produce high quality data that is comparable to laboratory experiment (Gupta et al., 2021; Peer et al., 2022). Recruiting from two platforms is important for my research as it allows me to cross check my findings and increases the reliability of my results. In addition, both platforms strictly forbid researchers from asking participants for their identifiable information. This enforced anonymity protects participants' privacy and is especially important for research on sensitive topics as it minimizes experimenter demand effects and social desirability bias. Finally, it is worth pointing out that recent evidence shows that the size of experimenter demand effect in online survey experiments is small. The knowledge of the experiment's purpose has no detectable effect on participants' behaviors (De Quidt et al., 2018; Mummolo & Peterson, 2019).

To maximize statistical power within my budget constraint, I only recruited subjects who met all the following conditions:

1. Identified themselves as Republican or Democrat.
2. Identified themselves as male.

[40] See Appendix A.2.3.9 for the survey script.

3. Currently resides in the United States.
4. Has a Minimum approval rate of 95%.[41]

I target both Republicans and Democrats for two reasons. First, It is well established that Americans' attitudes about race and gender are divided by partisanship (for example, (Doherty et al., 2019; Horowitz et al., 2017)). Recent survey evidence shows that while the majority of Democrats perceive establishing a path to legal status for immigrants as the top priority in immigration policy, the majority Republicans prioritize increasing border security and deportations of illegal immigrants.[42] Moreover, political speeches on immigration has become increasingly polarized. Computational analysis of US congressional speeches and presidential communications shows that Republicans are more likely to frame immigration in negative terms such as "crime" and "threats," while Democrats tend to use more positive framing such as "family" and "contribution" (Card et al., 2022). Thus, it is very likely that people who identify as Democrats and people who identify as Republicans will have very different views about immigration and immigrants. Since immigrant is the population that was targeted in the shooting on the news story, Stratifying the recruitment and the randomization by political affiliation allows me to test *Hypothesis 4*, i.e., whether the emphasis of the media coverage on the shooter and the shooter's ideology has a differential impact on the audience based on ex-ante biases towards the victimized group.

I recruit male subjects because men are far more likely to commit hate crimes than women. According to the FBI, the most common hate crime offense are vandalism, simple assault, and aggravated assault.[43] While the FBI does not publish hate crime statistics by sex, it is evident that men commits these offenses far more often than women. Men account for 76.5% of arrests for aggravate assault, 70.6% of arrests for other assaults, and 76.8% of arrests for vandalism.[44] This is confirmed by Lantz (2022), who shows that only 16% of the offenders in hate crimes are female-only. Although it is possible that women express more hatred than men, data suggests they do so in a less violent and extreme manner. In addition, since gender is not the focus of my

[41] A user's approval rate is the number of approved submissions divided by the number of total submissions.
[42] For example, click here to see a report by Pew Research Center.
[43] See FBI's 2020 hate crime statistics here.
[44] See FBI's latest release here.

study, recruiting only male subjects allows me to eliminate gender heterogeneity and increases statistical power. The experimental procedure including the inclusion restriction was preregistered on AsPredicted (#74996).

1.3.6 Estimation Strategy

I tests the effects of the information treatments on subjects' outcome measures by estimating the following equation using OLS on the full sample:

$$Y_i = \alpha + \beta_1 T_{1i} + \beta_2 T_{3i} + \beta_3 T_{4i} + \delta X_i + \epsilon_i \tag{1.1}$$

Where Y_i is the 4 sets of outcome variables described in Section 1.3.4. T_{1i} is an indicator variable that equals 1 if subject i is assigned to the No Hate Treatment, 0 otherwise. Similarly, T_{3i} is an indicator variable for the Hate Ideology Treatment, T_{4i} is an indicator variable for the Hate Background Treatment. The Hate Treatment is the omitted group in regression analysis. X is a set of individual characteristics measured before the information treatment: age, education level, income level, an indicator variable that equals 1 if the subject is White or Caucasian, an index for fame-seeking personality constructed using the Big 5 modesty measure (Konstabel et al., 2012),[45] an index for political stance constructed using subjects' stated views on a list of controversial political issues as described in Section 1.3.4.5, and an indicator variable that equals 1 if the subject is recruited from CloudResearch. β_1 captures the extent to which people naturally react to hate-motivated mass shootings differently. β_2 and β_3 capture the impacts of media emphasis on the shooter's ideology and media emphasis on the shooter's background on viewers' attitudes.

In order to statistically test whether there are heterogeneous treatment effects by baseline bias toward immigrants as specified in *Hypothesis 3*, I estimate equation 1.1 separately for the Democrat sample and the Republican sample. In addition, I estimate equation 1.2 below, where I pool the two samples and include interactions between the treatment dummies and a dummy for political

[45]The literature on mass shootings points out that shooters often have a desire for fame and attention Bushman (2018); Langman (2018); Lankford and Madfis (2018b); Silva and Greene-Colozzi (2019).

affiliations.

$$
\begin{aligned}
Y_i = \alpha + \beta_1 T_{1i} + \beta_2 T_{3i} + \beta_3 T_{4i} \\
+ \beta_4 T_{1i} * Republican_i + \beta_5 T_{3i} * Republican_i + \beta_6 T_{4i} * Republican_i \\
+ \gamma Republican_i + \delta X_i + \epsilon_i
\end{aligned}
\tag{1.2}
$$

$Republican_i$ is a dummy that equals to 1 if subject i is from the Republican sample. In this specification, β_1, β_2, and β_3 capture the treatment effects on subjects from the Democrat sample, whom I hypothesize to hold less ex-ante bias towards immigrants. β_4, β_5, β_6 capture the differential impacts of treatments on subjects from the Republican sample, whom I hypothesize to hold stronger ex-ante bias.

Since I have three outcome measures for attitudes toward the shooter, one concern would be multiple hypothesis testing causing false rejection of true null hypotheses. To address this concern, I construct a standardized index of support using responses to all three questions. I first standardize the response to each question with respect to the mean and standard deviation of the Hate Treatment. I then construct an inverse covariance weighted index using Anderson (2008)'s method. In addition, I report the sharpened False Discovery Rate (FDR) q-values in the Appendix.

As a secondary analysis, I examine possible mechanisms as specified in Section 1.3.4.5. In addition to investigating whether there are heterogeneous treatment effects between the Democrat sample and the Republican sample, I examine whether there are heterogeneous treatment effects within each sub-sample by estimating the following equation.

$$
\begin{aligned}
Y_i = \alpha + \beta_1 T_{1i} + \beta_2 T_{3i} + \beta_3 T_{4i} \\
+ \beta_4 T_{1i} * Political_i + \beta_5 T_{3i} * Political_i \\
+ \beta_6 T_{4i} * Political_i + \gamma Political_i + \delta X_i + \epsilon_i
\end{aligned}
\tag{1.3}
$$

Where $Political_i$ is an index for subject i's political stance constructed using the subject's response to six controversial political issue questions as explained in Section 1.3.4.5. An answer that aligns with the Democratic Party ideology will be coded as a 0, an answer that aligns with the Republican Party ideology will be coded as a 1. To generate an index, I calculate the average of the responses to all six questions, weighted by the inverse covariance matrix as Anderson (2008). The index is a continuous variable between 0 and 1 where a higher value represents a more right-leaning political stance. If the news treatments are persuading people with no initial bias against immigrants to hold bias, I should expect to see stronger treatment effects for people who are closed to the center of the political distribution, i.e., right-leaning Democrats and left-leaning Republicans. If that is the case, the estimated β_5 and β_6 should be positive for the Democrat sample, and negative for the Republican sample.

To examine whether changes in the perception of social norm serve as the mechanism behind the treatment effects. I estimate the same specification as equation 1.1, except that the outcome variables are replaced with the elicited norm measures. I conduct this exercise using the pooled sample, as well as the two sub-samples separately.

1.4 Results

In this section, I present the results from my experiment. Section 1.4.1 reports descriptive statistics of my sample. Section 1.4.2 reports treatment effects. Section 1.4.3 reports results from secondary analyses to identify the mechanism. Section 1.4.4 reports results from several robustness checks.

1.4.1 Descriptive Statistics

I recruited 1600 subjects from Prolific between Fall 2021 to Summer 2022. I recruited 800 subjects from CloudResearch in Fall 2022.[46] I stratified the recruitment and treatment assignment by political affiliation such that there is an equal number of Democrats and Republicans in each treatment condition. The experiment was programmed in Qualtrics. Subjects were paid $2.67 for

[46]Appendix Table A.9 reports the number of subjects by treatment condition and recruiting platform.

completing the survey,[47] and had the chance to earn up to \$1.2 of bonus payment.[48] The average completion time is around 16 minutes. The actual hourly earning translates into around \$14/hour, which is considerably high for online surveys.[49] In terms of data quality, 83.58% of subjects answered both comprehension questions correctly, 86.38% subjects passed both attention checks.

Table 1.1 reports demographic characteristics of my sample. Panel A reports the demographics of the full sample. Panel B and Panel C report the demographics of the Democrat sample and the Republican sample respectively. The last column of the table reports the p-value of an ANOVA test against the null hypothesis that subjects across the four experimental treatment conditions are not jointly different from each other. Subjects on average are around 38 years old and college educated. Overall, the full sample and the two sub-samples are balanced across different treatment conditions except for two instances. I account for these imbalances in the empirical analysis by controlling for the imbalanced variables.

There are several differences in demographics between the Democrat sample and the Republican sample. Panel A of Appendix Table A.10 reports test statistics for the Democrat sample vs Republican sample. On average, subjects from the democrat sample are younger, have higher education, lower income, less likely to be white, less fame-seeking. Unsurprisingly, the political index shows that the Democrat sample is much more left-leaning compared to the Republican sample (0.08 vs 0.62, p-value=0).[50]

In addition, there are several differences in demographics between Prolific subjects and CloudResearch subjects. Panel B of Appendix Table A.10 shows that on average, subjects from the Prolific sample are younger, more educated, have higher income, and more left-leaning. To account for the difference in recruiting platform in the regression analysis, I control for a dummy that equals 1 if the subject is recruited from CloudResearch. One concern is subjects who are registered on CloudResearch could also be registered on Prolific. Thus, it is possible that the same person will

[47]The completion fee for subjects who participated before April 2022 is \$2.17. Prolific increased its minimum wage from \$6.5 to \$8 an hour in April 2022. Thus, the completion fee for subsequent experiments is increased to \$2.67.

[48]As explained in Section 1.3.4, the incentivized norm questions are unavailable for the first 200 subjects in each sample. The maximum bonus for the initial 200 subjects is \$0.4 from answering comprehension questions correctly.

[49]For example, the average earnings on MTurk is \$2 per hour (Hara et al., 2018).

[50]The distribution of political stance is shown in Appendix Figure A.1.

Table 1.1: Online Experiment - Summary Statistics and Balance Tests

	All	No Hate	Hate	Hate Ideology	Hate Background	ANOVA p-value
Panel A - Full Sample (N = 2400)						
Age	38.26	37.36	38.79	38.03	38.85	0.16
	(13.16)	(12.64)	(13.44)	(13.29)	(13.22)	
Education level	3.76	3.78	3.80	3.75	3.70	0.47
	(1.10)	(1.14)	(1.08)	(1.14)	(1.05)	
Income level	7.11	7.18	7.28	7.17	6.83	0.10*
	(3.36)	(3.42)	(3.35)	(3.38)	(3.26)	
White	0.74	0.73	0.75	0.74	0.74	0.88
	(0.44)	(0.44)	(0.43)	(0.44)	(0.44)	
Political index	0.36	0.35	0.36	0.36	0.36	0.80
	(0.35)	(0.34)	(0.35)	(0.35)	(0.35)	
Fame-seeking index	2.71	2.76	2.69	2.68	2.69	0.33
	(0.88)	(0.89)	(0.91)	(0.87)	(0.84)	
Observations	2,400	602	601	597	600	
Panel B - Democrat Sample (N = 1199)						
Age	37.81	36.90	38.86	36.95	38.54	0.12
	(12.95)	(13.06)	(13.33)	(12.44)	(12.88)	
Education level	3.82	3.85	3.77	3.73	3.92	0.15
	(1.08)	(1.12)	(1.06)	(1.10)	(1.05)	
Income level	6.75	6.77	6.98	6.68	6.56	0.48
	(3.33)	(3.44)	(3.35)	(3.33)	(3.21)	
White	0.69	0.67	0.71	0.67	0.70	0.59
	(0.46)	(0.47)	(0.45)	(0.47)	(0.46)	
Political index	0.08	0.07	0.08	0.09	0.07	0.66
	(0.14)	(0.12)	(0.15)	(0.15)	(0.15)	
Fame-seeking index	2.56	2.59	2.54	2.54	2.58	0.86
	(0.89)	(0.90)	(0.94)	(0.88)	(0.84)	
Observations	600	301	300	297	301	
Panel C - Republican Sample (N = 1201)						
Age	38.70	37.82	38.71	39.11	39.15	0.59
	(13.35)	(12.22)	(13.56)	(14.02)	(13.57)	
Education level	3.70	3.71	3.82	3.78	3.48	0.00***
	(1.12)	(1.15)	(1.11)	(1.19)	(1.01)	
Income level	7.48	7.59	7.57	7.65	7.09	0.15
	(3.35)	(3.36)	(3.34)	(3.36)	(3.31)	
White	0.79	0.78	0.78	0.80	0.79	0.92
	(0.41)	(0.41)	(0.41)	(0.40)	(0.41)	
Political index	0.64	0.62	0.64	0.64	0.64	0.51
	(0.26)	(0.27)	(0.26)	(0.27)	(0.25)	
Fame-seeking index	2.85	2.94	2.84	2.83	2.80	0.21
	(0.84)	(0.84)	(0.85)	(0.84)	(0.83)	
Observations	600	301	301	300	299	

Notes: A total of 2400 individuals participated in the online experiment. Subjects are recruited from Prolific and CloudResearch. This table reports the mean of each demographic variable across different treatment conditions. The corresponding standard deviation is reported in parentheses. Education level is a categorical variable ranging from Less than high school (1) to Doctorate (7). Income level is a categorical variable ranging from Less than $10,000 (1) to $15,000 or more (12). Political index ranges from 0 to 1, a higher value means more right-leaning. Fame-seeking index ranges from 1 to 5, a higher value means more fame-seeking.

participate in the experiment twice. To investigate this possibility, I additionally ask subjects from CloudResearch two questions in the end of the survey: 1) Are you registered on other recruiting platforms other than MTurk? 2) Have you seen this survey before on a different platform? While about 13% of CloudResearch subjects are also registered Prolific, only about 1% (10 subjects) say they saw the same survey before. Therefore, this concern is trivial.

1.4.2 Treatment Effects

In this subsection, I compare different contents of news coverage to causally identify the impact of news coverage of hate-motivated mass shootings. Section 1.4.2.1 examines the treatment effects on readers' interest. Section 1.4.2.2 examines the treatment effects on support for the shooter. Section 1.4.2.3 examines the treatment effects on support for the shooter's ideology, as measured by donations to an anti-immigrant organization. Section 1.4.2.4 examines the treatment effects on hatred, as measured by subjects' interest in white supremacy hate groups. As specified in my preregistration, I also investigate heterogeneous treatment effects by political affiliation.

1.4.2.1 Treatment Effects on Public Interest

I first test whether providing more information about the shooter increases viewers' interest in the shooting and the shooter. Panel A of Table 1.2 reports estimates from equation 1.1 using the full sample. The dependent variables are an indicator variable that equals 1 if the subject requests to receive access to the shooter's manifesto at the end of the survey, and an indicator variable that equals 1 if the subject requests to receive access to more information about the shooter's background. For each variable, I report estimates without controls in odd columns, and estimates with controls in even columns.

I start by comparing the information demand in the No Hate Treatment and the Hate Treatment. The two treatment conditions are identical, except that the Hate Treatment mentions the hateful nature of the shooting. In Section 1.2, I showed that the media has a tendency to focus on the shooter in hate-motivated mass shootings. Moreover, the online searching behavior after mass shootings reveals that people show much higher interest when the shooting is hate-motivated. One possible explanation is hate crimes simply draw more attention from people, regardless of what the media does. In which case subjects in the Hate Treatment should show higher information demand compared to subjects in the No Hate Treatment.

Estimates in Panel A of Table 1.2 shows the contrary: subjects' information demand significantly decreased when they learn that the shooting is hate-motivated. Compared with subjects in the Hate Treatment, subjects in the No Hate Treatment are 5.7 percentage points more likely to request information about the shooter's ideology, and 5.9 percentage points more likely to request information about the shooter's background. This pattern shows that people are not naturally more interested in hate-motivated mass shootings. If anything, the decrease in information demand in the Hate Treatment suggests that people have a natural distaste for hate crimes. This implies that the difference in viewer's interest displayed in the observational data is likely to be caused by media coverage instead of viewer's natural preferences, either through the difference in content, or the difference in intensity/duration.

Table 1.2: Treatment Effects on Interest

	A: Full Sample				B: Democrat				C: Republican			
	Manifesto		Background		Manifesto		Background		Manifesto		Background	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
No Hate (T1)	0.057**	0.050*	0.059**	0.055*	0.055	0.047	0.058	0.056	0.060	0.057	0.060	0.060
	(0.028)	(0.028)	(0.029)	(0.029)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Hate Ideology (T3)	-0.083***	-0.086***	-0.057**	-0.058**	-0.016	-0.026	-0.001	-0.004	-0.148***	-0.146***	-0.111***	-0.111***
	(0.028)	(0.028)	(0.029)	(0.029)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Hate Background (T4)	-0.034	-0.036	-0.126***	-0.125***	-0.031	-0.030	-0.135***	-0.130***	-0.037	-0.038	-0.117***	-0.114***
	(0.028)	(0.028)	(0.029)	(0.029)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.041)
Control Mean	0.436	0.436	0.557	0.557	0.407	0.407	0.563	0.563	0.465	0.465	0.551	0.551
Observations	2,400	2,400	2,400	2,400	1,199	1,199	1,199	1,199	1,201	1,201	1,201	1,201
R-squared	0.011	0.031	0.019	0.024	0.004	0.032	0.020	0.033	0.023	0.039	0.023	0.027
T1=T3 p-value	0.000***	0.000***	0.000***	0.000***	0.077*	0.066*	0.145	0.141	0.000***	0.000***	0.000***	0.000***
T1=T4 p-value	0.001***	0.002***	0.000***	0.000***	0.031**	0.052*	0.000***	0.000***	0.016**	0.019**	0.000***	0.000***
T3=T4 p-value	0.090*	0.075*	0.016**	0.019**	0.706	0.921	0.001***	0.002***	0.006***	0.007***	0.897	0.937
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variable in columns that are labeled Manifesto is an indicator variable that equals 1 if the subject requested to be shown the shooter's manifesto. The dependent variable in columns that are labeled Background is an indicator variable that equals 1 if the subject requested to be shown the shooter's Background. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

I then investigate whether differences in the content of news coverage causes difference in public interest. Recall that compared to subjects in the Hate Treatment, subjects in the Hate Ideology Treatment saw additional information on the shooter's ideology, and subjects in the Hate Background Treatment saw additional information on the shooter's background. Estimates in Panel A of Table 1.2 show that subjects who are assigned to the more informative treatment conditions show significantly less interest in receiving more information. In fact, the decrease in information demand is the largest when subjects read the news story that covers that information. Subjects in the Hate Ideology Treatment are 8.25 percentage points less likely to request information about the shooter's ideology (compared to a mean of 43.59 percent in the Hate Treatment). Subjects in the Hate Background Treatment are 12.57 percentage points less likely to request information about the shooter's background (compared to a mean of 55.74 percent in the Hate Treatment). The results are robust to the inclusion of control variables. These findings rule out the possibility that people's high interest in hate-motivated mass shootings is driven by the content of news coverage. Instead, public interest is more likely to be driven by media's tendency to spend more time covering hate-motivated mass shootings as shown in section 1.2.2.1.

Panel B and Panel C of Table 1.2 report estimates separately for the Democrat sample and the Republican sample. In the baseline treatment (Hate), Democrat subjects show slightly lower interest in the shooter's manifesto (0.407 vs 0.465, p-value=0.149). The signs of the coefficients on the treatment dummies are largely consistent with that of the full sample. Subjects from both samples show lower interest in the more informative treatment conditions.

1.4.2.2 Treatment Effects on Support for the Shooter

I test whether emphasizing the shooter's hate ideology (T3) or the shooter's identity and background (T4) increases support for the shooter. As described in section 1.3.4.2, I ask subjects three questions to measure their attitude towards the shooter: 1) How much they admire the shooter's courage, 2) How much they believe the shooter's action can be justified, 3) What they think the sentencing for the shooter should be. To construct an overall measure of support, I use Anderson (2008)'s method and the responses to all three questions to generate an index of support standardized around the control mean as described in Section 3.3.5. Thus, the index is displayed in standard deviations from the mean of the Hate Treatment.

In section 1.2, I showed that media has a tendency to report stories about the ideology and background of the shooter in hate-motivated mass shootings. Moreover, there is an increase in the number of hate crimes against the same population that was targeted in the shooting. If the increase in hate crime is because of the media's focus on the shooter that caused some people to admire, worship, or even copying the shooter, I should expect subjects in the Hate Ideology Treatment and the Hate Background Treatment to show higher support for the shooter as measured by the index.

Appendix Table A.11 reports estimates from equation 1.1 for all three attitudinal measures and shows that they move in the same direction.[51] Panel A of Table 1.3 reports estimates where the outcome variable is the support index using the same specification and the full sample. All the coefficients on treatment dummies are positive and significant. Across all treatments, subjects in the No Hate Treatment who are not informed of the shooting's hateful nature exhibit the strongest support for the shooter. Compared to subjects in the Hate Treatment who learns that the shooting is possibly a hate crime, subjects in the No Hate Treatment significantly increase their support for the shooter by about 0.22 standard deviations. The decrease in support when subjects learn the hateful nature of the shooting is consistent with the decrease in interest shown in the last section. However, this decrease is partially mitigated when subjects are shown additional information about the shooter's ideology in the Hate Ideology Treatment and the shooter's background information in the Hate Background Treatment. In fact, subjects in the Hate Ideology Treatment and the Hate Background Treatment show significantly higher support for the shooter compared to subjects in the Hate Treatment. Knowing either the shooter's ideology or the shooter's background make subjects more supportive by about 0.12 standard deviations. The results are robust to the inclusion of control variables. These findings provide support for *Hypothesis 3*, the news emphasis on the

[51] Panel B and Panel C of Appendix Table A.11 show regressions on all three attitudinal measures separately for two sub-samples. On the aggregate level, Republican subjects show higher admiration for the shooter, higher justification for the shooter's action, but choose harsher sentencing options for the shooter (A t-test for mean comparison returns a p-value of 0 for all three variables). This pattern is consistent with Republican's tougher attitude against immigrants, and Democrat's stronger opposition to using the death penalty (For example, see This report).

shooter's ideology and background increases support for the shooter.

Table 1.3: Treatment Effects on Support for the Shooter

	Index of support for the shooter					
	A: Full Sample		B: Democrat		C: Republican	
	(1)	(2)	(3)	(4)	(5)	(6)
No Hate (T1)	0.217***	0.194***	0.189***	0.168**	0.246***	0.234***
	(0.050)	(0.048)	(0.070)	(0.066)	(0.070)	(0.067)
Hate Ideology (T3)	0.122**	0.117**	0.057	0.035	0.174**	0.188***
	(0.050)	(0.048)	(0.070)	(0.067)	(0.071)	(0.067)
Hate Background (T4)	0.121**	0.126***	0.280***	0.265***	0.001	0.043
	(0.050)	(0.048)	(0.070)	(0.066)	(0.071)	(0.068)
Control Mean	0.000	0.000	0.000	0.000	0.000	0.000
Observations	2,400	2,400	1,199	1,199	1,201	1,201
R-squared	0.008	0.071	0.016	0.121	0.015	0.114
T1=T3 p-value	0.055*	0.109	0.059*	0.045**	0.307	0.494
T1=T4 p-value	0.054*	0.163	0.191	0.143	0.001***	0.005***
T3=T4 p-value	0.993	0.836	0.001***	0.001***	0.015**	0.032**
Control Variables	No	Yes	No	Yes	No	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variable in each panel is an index measuring support for the shooter that is standardized around the control mean (Hate Treatment) of each sample. The estimates are represented in standard deviations from the control mean. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Panel B and Panel C of Table 1.3 show the regression on the standardized support index for two sub-samples separately. For both the Democrat sample and the Republican sample, the coefficient on No Hate is positive and significant. Thus, the decrease in support for the shooter when subjects learn the shooting is hate-motivated is true for regardless of ex-ante bias towards the victims. However, the coefficients on Hate Ideology and Hate Background show substantial heterogeneous treatment effects. The Hate Ideology Treatment significantly increases Republican subjects' support for the shooter by 0.188 standard deviations, but has no effect on Democrat subjects. This finding is consistent with Demszky et al. (2019), who uses Twitter data to show that Republicans are more likely to focus on the shooter's ideology when discussing a mass shooting. In contrast, the Hate Background Treatment significantly increases Democrat subjects' support for the shooter by 0.265 standard deviations, but has no effect on Republican subjects. These results are confirmed by the analysis shown in column 5 and column 6 of Table A.12, which reports estimates from equation 1.2 that includes interaction terms between the treatment dummies and party dummy. In fact, the level increase is almost identical for Democrat subjects in the Hate Background Treatment and Republican subjects in the Hate Ideology Treatment. The differential treatment effects provide support for *Hypothesis 4*.

It is worth pointing out that the overall level of support for the shooter shown in the experiment is quite low. The majority of subjects strongly disagree that they admire the shooter's courage, and strongly disagree that the shooter's action can be justified. Thus, one might question the magnitude of effect given the low baseline level. I discuss this concern in more detail in Section 1.4.5.

1.4.2.3 Treatment Effects on Support for the Shooter's Ideology

I now investigate whether emphasizing the shooter's hate ideology (T3) or the shooter's identity and background (T4) increases support for the shooter's ideology, measured by donation to an anti-immigrant organization. In the last section, I showed that subjects show higher support for the shooter when they read a news story that mentions either the shooter's ideology or background. However, it is not clear whether this increase in support actually translates into action. It is possible that people dislikes the shooter's hateful ideology but simply idolize the shooter as a person for other reasons. If there is indeed an increase in support for the shooter's ideology, subjects in the Hate Ideology Treatment and the Hate Background Treatment should be more likely to authorize the donation to the anti-immigrant organization. Panel A of Table 1.4 reports estimates from equation 1.1 using the full sample. The dependent variable in column 1 and column 2 is an indicator variable that equals 1 if the subject authorized an 1$ donation to the anti-immigrant organization. As a robustness check, I also similarly examine how the treatments affect donation to the pro-immigration organization in column 3 and column 4. First, at the baseline level, the donation rate to the pro-immigrant organization (63.8%) is considerably higher than the donation rate to the anti-immigrant organization (17.8%). This is consistent with the difference in the framing used when describing the organizations: FAIR (anti-immigrant) is described as an organization with ties to white supremacy and racism, while AIC (pro-immigrant) is described using neutral language, as an organization working to improve the US immigration system. Second, in contrast to the increase in support for the shooter shown in the last section, there's no difference in the donation rate to either organization across all treatment conditions. None of the coefficients on treatment dummies are statistically different from zero. Therefore, there is no evidence that the content of media coverage affected the support for the shooter's anti-immigrant ideology.

While the estimates from the full sample does not support my hypotheses, the null results could be due to two reasons. First, it is possible that my news treatments are not strong enough to generate changes in behavior. As shown in Section 1.2.2, the intensity of media coverage on hate-motivated mass shootings is high. It is likely that news viewers are repeatedly exposed to stories about the shooter's ideology and background. Thus, a one time exposure in my experiment might not be sufficient. Second, it is also possible that there is heterogeneity in treatment effects as suggested in *Hypothesis 4*, and this heterogeneity caused the overall lack of effects.

Table 1.4: Treatment Effects on Support for the Shooter's Ideology

	A: Full Sample				B: Democrat				C: Republican			
	Donation anti-immigrant		Donation pro-immigrant		Donation anti-immigrant		Donation pro-immigrant		Donation anti-immigrant		Donation pro-immigrant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
No Hate (T1)	-0.006	-0.007	-0.060	-0.053	0.028	0.021	-0.077	-0.067	-0.033	-0.037	-0.016	-0.035
	(0.028)	(0.027)	(0.050)	(0.044)	(0.035)	(0.034)	(0.062)	(0.061)	(0.042)	(0.042)	(0.070)	(0.065)
Hate Ideology (T3)	-0.016	-0.016	-0.048	-0.041	0.020	0.013	-0.086	-0.082	-0.046	-0.045	0.017	-0.004
	(0.027)	(0.027)	(0.052)	(0.046)	(0.034)	(0.033)	(0.065)	(0.063)	(0.041)	(0.041)	(0.073)	(0.067)
Hate Background (T4)	-0.016	-0.016	-0.038	-0.034	0.071**	0.069**	0.047	0.039	-0.101**	-0.093**	-0.099	-0.095
	(0.027)	(0.026)	(0.053)	(0.047)	(0.034)	(0.033)	(0.065)	(0.064)	(0.041)	(0.041)	(0.075)	(0.070)
Control Mean	0.198	0.198	0.623	0.623	0.114	0.114	0.778	0.778	0.278	0.278	0.457	0.457
Observations	1,665	1,665	735	735	842	842	357	357	823	823	378	378
R-squared	0.000	0.045	0.002	0.211	0.006	0.088	0.015	0.087	0.008	0.034	0.007	0.184
T1=T3 p-value	0.718	0.746	0.819	0.787	0.818	0.817	0.882	0.807	0.760	0.853	0.640	0.625
T1=T4 p-value	0.693	0.753	0.666	0.678	0.205	0.141	0.059*	0.098*	0.104	0.179	0.249	0.371
T3=T4 p-value	0.975	0.991	0.843	0.885	0.128	0.085*	0.051*	0.070*	0.180	0.237	0.121	0.189
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variable in columns that are labeled "Donation anti-immigrant" is an indicator variable that equals 1 if the subject authorized the 1$ donation to the anti-immigration organization. The dependent variable in columns that are labeled "Donation pro-immigrant" is an indicator variable that equals 1 if the subject authorized the 1$ donation to the pro-immigration organization. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

I investigate heterogeneous treatment effects by estimating equation 1.1 separately for the Democrat sample and Republican sample. Estimates are reported in Panel B and Panel C of Table 1.4. First, consistent with the partisan and ideological differences on immigration policy, Republican subjects are significantly more likely to donate to the anti-immigrant organization (0.278 vs 0.114 in the Hate Treatment, p-value=0), and significantly less likely to donate to the pro-immigrant organization compared to Democrat subjects (0.457 vs 0.778 in the Hate Treatment, p-value=0). The coefficient on the Hate Background Treatment shows substantial heterogeneous treatment effects by political affiliation. Democrat subjects significantly increase their likelihood of donating to the anti-immigrant organization by about 7 percentage points (61% increase) when they are given background information of the shooter. This pattern is consistent with the increase in support for the shooter as shown in the last subsection. In contrast, the same treatment condition significantly decreases Republican subjects' likelihood of donating to the anti-immigrant organization by 9.3 percentage points (33% decrease). These results are confirmed by the estimates shown in column 7 and column 8 of Table A.12. When the news story contains information shooter's background, subjects from the Democrat sample who are generally characterized by their friendly attitude towards immigrants, become more anti-immigrants, subjects from the Republican sample who are generally characterized by their unfavorable attitude towards immigrants, decreases their support for anti-immigrant ideology. The contrasting treatment effects on Democrats and Republicans support *Hypothesis 4*.

In contrast, the likelihood of donating to the pro-immigrant organization is not statistically different across different treatment conditions for either Democrat subjects or Republican subjects. While I expect the information treatments to affect the support for pro-immigrant in opposite directions, there are several factors that can explain the null effect. First, by design, only 25% of the subjects are given the opportunity to donate to the pro-immigrant organization. Thus, the small sample size might not be statistically sufficient to detect any variation. Second, people might not perceive AIC (pro-immigrant) as the opposite of FAIR (anti-immigrant). By design, FAIR is described in strong language to make sure subjects understand the organization's stance. In compassion, AIC is described in neutral language as an organization whose goal is to improve the immigration system of United States. Thus, it is possible that subjects with anti-immigrant sentiment would nevertheless donate to AIC. Third, the results might be caused by social desirability bias. In particular, donating to a pro-immigrant organization may be perceived as especially desirable. Thus, subjects would authorize the donation regardless of the news story they see. I discuss issues related to social desirability bias in more detail in Section 1.4.4.

1.4.2.4 Treatment Effects on Interest in Hate Group

To test if emphasis on the shooter's ideology or the shooter's identity and background increases hatred, I examine the treatment effect on subjects' interest in a white-supremacy hate group named Stormfront. In the survey, subjects are shown a description of the hate group, and are asked if they would like to receive links to the hate group's website. I examine two outcomes: 1) whether a subject requested the links, 2) whether a subject actually clicked on the links after requesting them. First, it is worth pointing out that the fraction of subjects who expressed interest is extremely low. At the aggregate level, only 10.67% of the subjects requested the links for accessing Stormfront's website, and only 1.04% of the subjects clicked on the links. The low volume of requests and clicks shows that the interest in hate group is unlikely to be driven by pure curiosity. Therefore, the variation in interest is at least partially capturing the change in hatred. However, consistent with the null result in donation rate, subjects' interest does not vary by treatment conditions. Table 1.5 presents estimates from equation 1.1. There is no statistically significant difference in the percentage of subjects who requested the links or the percentage of subjects who clicked on the links across different treatment conditions. This result does not support my hypotheses.

Panel B and Panel C of Table 1.5 report the estimates from equation 1.1 separately for the Democrat sample and the Republican sample. First, consistent with the higher support for the shooter and the higher donation rate to the anti-immigrant organization shown in previous subsections, Republican subjects are more likely to request the links for access at the baseline (0.13 vs 0.067, p-value=0.01). Second, for both samples, the percentage of subjects who requested the links does not seem to vary much across different treatments.

Table 1.5: Treatment Effects on Interest in Hate Group

	A: Full Sample				B: Democrat				C: Republican			
	Links requested		Links clicked		Links requested		Links clicked		Links requested		Links clicked	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
No Hate (T1)	0.015	0.009	-0.007	-0.007	0.016	0.012	-0.003	-0.003	0.013	0.012	-0.010	-0.011
	(0.018)	(0.018)	(0.006)	(0.006)	(0.023)	(0.022)	(0.007)	(0.007)	(0.027)	(0.027)	(0.009)	(0.009)
Hate Ideology (T3)	0.007	0.006	0.007	0.007	0.018	0.013	0.003	0.004	-0.003	0.002	0.010	0.010
	(0.018)	(0.018)	(0.006)	(0.006)	(0.023)	(0.022)	(0.007)	(0.007)	(0.027)	(0.027)	(0.009)	(0.009)
Hate Background (T4)	0.012	0.012	-0.005	-0.005	0.040*	0.040*	0.003	0.004	-0.016	-0.007	-0.013	-0.013
	(0.018)	(0.018)	(0.006)	(0.006)	(0.023)	(0.022)	(0.007)	(0.007)	(0.027)	(0.027)	(0.009)	(0.009)
Control Mean	0.098	0.098	0.012	0.012	0.067	0.067	0.007	0.007	0.130	0.130	0.017	0.017
Observations	2,400	2,400	2,400	2,400	1,199	1,199	1,199	1,199	1,201	1,201	1,201	1,201
R-squared	0.000	0.039	0.003	0.008	0.003	0.058	0.001	0.012	0.001	0.036	0.006	0.012
T1=T3 p-value	0.677	0.854	0.022**	0.020**	0.961	0.965	0.338	0.360	0.553	0.705	0.032**	0.024**
T1=T4 p-value	0.868	0.862	0.774	0.732	0.307	0.216	0.346	0.370	0.286	0.482	0.724	0.790
T3=T4 p-value	0.802	0.721	0.045**	0.047**	0.332	0.236	0.985	0.983	0.636	0.745	0.013**	0.012**
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variable in columns that are labeled Requested is an indicator variable that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group. The dependent variable in columns that are labeled Clicked is an indicator variable that equals 1 if the subject clicked on the provided links. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

There is some evidence that the Hate Background Treatment increases Democrat subjects' likelihood of requesting the links by 4 percentage points (60% increase, p-value=0.076). This pattern is consistent with the increase in support for the shooter and the shooter's ideology shown in previous subsections. When Democrat subjects learn about the shooter's background story, they increase their support for the shooter measured by admiration, justification, and sentencing for the shooter, they increase their support for the shooter's ideology measured by a $1 donation to an anti-immigrant organization, and they show higher interest in white supremacy hate groups. However, my sample size is not large enough for a more precise estimate. Moreover, the same treatment does not change Democrat subjects' likelihood of clicking on the links. Column 9 to column 12 of Table A.12 presents estimates from the fully saturated model including interaction terms between treatment dummies and party dummy. None of the coefficients are statistically significant after controlling for demographic variables. Therefore, the estimates presented in Table 1.5 are susceptible to bias and should be interpreted with caution.

1.4.3 Secondary Analyses

In the last section, I showed that knowing the shooter's background story increases Democrat subjects' support for the shooter, support for the shooter's ideology, and interest in white-supremacy hate groups (suggestive evidence). In contrast, knowing the shooter's ideology increases Republican subjects' support for the shooter. In this section, I conduct several secondary analyses as specified in Section 3.3.5.

First, I examine whether there are heterogeneous treatment effects within each sub-sample by estimating equation 1.3. If the news treatments persuaded people to develop bias against immigrants, then the treatment effects should be driven by subjects who are positioned near the middle of the political spectrum, i.e., right-leaning Democrats and left-leaning Republicans. This population is likely to be neutral toward immigrants and thus may be especially subjective to media's influence. Table 1.6 report the estimates. Panel A reports estimates for the Democrat sample. The coefficient on the interaction term between the political index and the dummy for the Hate Background Treatment is positive and significant. This shows that the observed treatment effects of knowing the shooter's background story on Democrat subjects are entirely driven by the right leaning subjects within the Democrat sample. The magnitude of the point estimates are large. For example, for Democrat subjects in the Hate Background Treatment, a 0.1 increase (about 0.7 standard deviations) in the political index will increase their likelihood of donating to the anti-immigrant organization by 6.7 percentage points. In comparison, there is little heterogeneity within the Republican sample. The interaction terms between the political index and the treatment dummies are largely insignificant. One concern is the political index is constructed using the subject's view on six diverse political issue questions and does not necessarily capture the subject's attitude toward immigrants. To address this concern, I measure subjects' attitude toward immigrants using only two political issue questions related to immigration. The results shown in Appendix Table A.13 are largely consistent with Table 1.6. These findings provide support that the news treatments are persuading right-leaning Democrats to be develop hatred toward immigrants. This is also consistent with Song (2021), who shows that racial progressive content on Twitter makes racial moderates more progressive, but has little effect on racial progressives and conservatives.

Second, I examine social norms as a possible mechanism for the treatment effects. As described in Section 1.3.4, I elicit subjects' perception of social norms on the support for the shooter. I incentivize subjects to guess what they think is the option that was chosen the most by previous survey participants when asked: 1) How much they admire the shooter's courage, 2) How much they believe the shooter's action can be justified, 3) What they think the sentencing for the shooter should be. Subjects are rewarded \$0.2 for each correct guess. To construct an overall measure of norm perception, I standardize the response to each of the three norm questions around the control mean and use Anderson (2008)'s method to generate an index of norms. Similarly, after subjects makes the decision to donate, I incentivize them to guess what percentage of previous subjects authorized the donation. Table 1.7 reports estimates from equation 1.1 where the dependent variables are the aforementioned norm variables. Results show that subjects from both parties overestimate the social norms. For example, Democrat subjects in the Hate Treatment guessed 24% of previous survey participants authorized the donation when the true percentage is only 11%. However, the

Table 1.6: Heterogeneous Treatment Effects within Sub-samples

	A: Democrat sample				B: Republican sample			
	Index Support (1)	Donation anti-immigrant (2)	Links requested (3)	Links clicked (4)	Index Support (5)	Donation anti-immigrant (6)	Links requested (7)	Links clicked (8)
No Hate (T1)	0.104	0.026	0.014	-0.005	0.661***	0.036	0.079	-0.023
	(0.075)	(0.039)	(0.025)	(0.008)	(0.175)	(0.109)	(0.070)	(0.024)
Hate Ideology (T3)	0.013	0.005	0.008	0.002	0.104	0.060	0.001	0.014
	(0.075)	(0.038)	(0.025)	(0.008)	(0.176)	(0.109)	(0.071)	(0.025)
Hate Background (T4)	0.072	0.020	0.006	-0.013	0.049	-0.140	0.058	-0.028
	(0.073)	(0.037)	(0.025)	(0.008)	(0.183)	(0.112)	(0.074)	(0.026)
Political Index (PI)	0.493	0.247	0.081	-0.031	-0.089	0.150	0.067	0.001
	(0.323)	(0.174)	(0.109)	(0.034)	(0.184)	(0.111)	(0.074)	(0.026)
No Hate * PI	0.847*	-0.074	-0.028	0.024	-0.686***	-0.116	-0.107	0.020
	(0.496)	(0.260)	(0.168)	(0.053)	(0.256)	(0.160)	(0.103)	(0.036)
Hate Ideology * PI	0.339	0.128	0.076	0.034	0.132	-0.162	0.002	-0.005
	(0.442)	(0.224)	(0.150)	(0.047)	(0.254)	(0.157)	(0.102)	(0.035)
Hate Background * PI	2.605***	0.671***	0.448***	0.225***	-0.011	0.071	-0.102	0.023
	(0.439)	(0.229)	(0.149)	(0.047)	(0.264)	(0.162)	(0.106)	(0.037)
Control Mean	0.000	0.114	0.067	0.007	0.001	0.278	0.130	0.017
Observations	1,199	842	1,199	1,199	1,201	823	1,201	1,201
R-squared	0.151	0.102	0.068	0.036	0.124	0.037	0.037	0.013
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the Democrat sample. Panel B shows the results for the Republican sample. The dependent variables in each panel from left to right are: (1) a standardized index measuring support for the shooter, (2) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (3) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (4) a dummy that equals 1 if the subject clicked on the provided links about the hate group. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, a dummy for the Hate Background Treatment, Political Index (PI), and interaction terms between the treatment dummies and the Political Index. The Hate Treatment is the omitted group. Political Index (PI) is an index that measures the subject's political stance, a higher value means more right-leaning. Control variables include age, income, education, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and in indicator variable that equals 1 if the subject is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

movement of subjects' norm perceptions between treatments is closely aligned with the movement of their actual behaviors. Democrat subjects who are shown the background story of the shooter and Republican subjects who are shown the ideology of the shooter significantly increase their their perceived popularity of the shooter. This shows that news coverage can shift subjects' perceptions of social norms. The change in acceptability of supporting the shooter and the shooter's ideology might explain the observed treatment effects.

1.4.4 Robustness Checks

In this subsection, I address several concerns that may threaten the validity of my results, including experimenter demand effect, inattention, recall bias, social desirability bias, and multiple

Table 1.7: Social Norm as a Mechanism

	A: Full Sample		B: Democrat		C: Republican	
	Index support norm	Donation anti-immigrant norm	Index support norm	Donation anti-immigrant norm	Index support norm	Donation anti-immigrant norm
	(1)	(2)	(3)	(4)	(5)	(6)
No Hate (T1)	0.122**	0.000	0.042	0.005	0.189***	-0.007
	(0.048)	(0.017)	(0.067)	(0.023)	(0.065)	(0.024)
Hate Ideology (T3)	0.096**	-0.002	0.032	0.010	0.146**	-0.016
	(0.048)	(0.017)	(0.067)	(0.023)	(0.065)	(0.023)
Hate Background (T4)	0.103**	-0.016	0.262***	0.050**	-0.003	-0.078***
	(0.048)	(0.017)	(0.067)	(0.023)	(0.065)	(0.023)
Control Mean	0.000	0.275	-0.000	0.240	-0.000	0.308
Observations	2,400	1,665	1,199	842	1,201	823
R-squared	0.059	0.034	0.129	0.089	0.087	0.071
T1=T3 p-value	0.586	0.915	0.888	0.835	0.501	0.703
T1=T4 p-value	0.680	0.347	0.001	0.052	0.003	0.003
T3=T4 p-value	0.894	0.397	0.001	0.079	0.023	0.007
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the Democrat sample. Panel B shows the results for the Republican sample. The dependent variables in each panel from left to right are: (1) a standardized index measuring the social norm of support for the shooter, (2) subjects' guess on the percentage of previous survey participants who authorized the $1 donation to the anti-immigrant organization. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

hypothesis testing.

1.4.4.1 Experimenter Demand Effect

One concern is that my results might be driven by experimenter demand effect. For example, subjects might correctly guessed the true purpose of the experiment and thus change their behavior. In addition, since my outcome measures are particularly sensitive, my results could be especially vulnerable to bias. There are several features of my design that help address this concern. First, my experiments were conducted online. The procedure is completely anonymous. I do not have any access to the participants' identifying information. The anonymous environment should help alleviate experimenter demand effect. In addition, recent literature shows that demand effects observed in experiments and surveys are modest at best (De Quidt et al., 2018). Moreover, online survey experiments are particularly resistant to demand effects (Mummolo & Peterson, 2019).

Second, as described in Section 1.3, my survey implements extra layers of randomization intended to cover the true purpose of my study and reduce demand effects. One might also be concerned that participants are self-selecting into the study. However, the selection bias should be minimal. Appendix Figure A.2 shows a screenshot of the study's advertisement on Prolific, which describes the study in neutral language. Moreover, more than 90% of the participants who started the study completed the study.

Third, I ask participants an open-ended question in the end of the survey "If you had to guess, what do you think is the purpose of our study?" If the subject's response contains words related to any of the following roots: 1)immigrant, 2)race, 3)hate, I identify that subject as correctly guessed the purpose of the study and may be prone to bias. Appendix Table A.14 reports the average percentage of subjects who guessed correctly in each treatment condition. Column 1 shows that 18% of subjects overall are able to correctly guess the purpose of my study. The high proportion could be due to the diverse keywords I used for identification. Column 2 to column 4 show that only 8%,6%, and 6% of subjects guessed correctly if I restrict keywords to those related to only one of the three roots. To examine whether there is significant difference in experimenter demand effect across different treatments conditions, I estimate equation 1.1 where the dependent variable is a dummy that equals 1 if subject correctly guessed the purpose of the experiment. Results are reported in Appendix Table A.15. Overall, subjects in the Hate Treatment are significantly more accurate than subjects in the No Hate Treatment. While there is little difference between the Hate Treatment and the Hate Background Treatment, subjects in the Hate Ideology are 4 percentage points more likely to correctly guess the experiment's purpose (p-value=0.066). I then examine whether my results still hold if I exclude the 18% of subjects who correctly guess the purpose of my study from regressions. The estimates reported in Table 1.8 are very similar in magnitude compared to the main estimates in Section 1.4.2. More importantly, the coefficients remain statistically significant. Thus, my findings are unlikely to be driven by experimenter demand effect.

Table 1.8: Treatment Effects Excluding Subjects who Guessed the Experiment's Purpose

	A: Full Sample				B: Democrat				C: Republican			
	Index support (1)	Donation anti-immigrant (2)	Links requested (3)	Links clicked (4)	Index support (5)	Donation anti-immigrant (6)	Links requested (7)	Links clicked (8)	Index support (9)	Donation anti-immigrant (10)	Links requested (11)	Links clicked (12)
No Hate (T1)	0.194***	-0.011	0.011	-0.005	0.174**	0.039	0.014	-0.004	0.227***	-0.059	0.012	-0.005
	(0.052)	(0.030)	(0.019)	(0.006)	(0.072)	(0.038)	(0.025)	(0.008)	(0.073)	(0.045)	(0.030)	(0.010)
Hate Ideology (T3)	0.147***	-0.006	0.019	0.010	0.028	0.035	0.022	0.000	0.251***	-0.040	0.018	0.021*
	(0.055)	(0.030)	(0.020)	(0.007)	(0.075)	(0.038)	(0.026)	(0.008)	(0.078)	(0.047)	(0.032)	(0.011)
Hate Background (T4)	0.143***	-0.020	0.018	-0.002	0.280***	0.082**	0.040	0.004	0.038	-0.119**	-0.001	-0.007
	(0.054)	(0.030)	(0.020)	(0.007)	(0.075)	(0.038)	(0.026)	(0.008)	(0.076)	(0.046)	(0.031)	(0.011)
Control Mean	0.001	0.206	0.101	0.011	0.000	0.112	0.069	0.009	0.001	0.289	0.131	0.012
Observations	1,960	1,368	1,960	1,960	991	704	991	991	969	664	969	969
R-squared	0.076	0.043	0.035	0.007	0.143	0.098	0.055	0.014	0.126	0.032	0.034	0.014
T1=T3 p-value	0.374	0.860	0.691	0.024	0.043	0.918	0.737	0.586	0.752	0.693	0.849	0.013
T1=T4 p-value	0.326	0.761	0.715	0.608	0.135	0.229	0.290	0.323	0.011	0.194	0.680	0.842
T3=T4 p-value	0.941	0.638	0.971	0.090	0.001	0.202	0.490	0.674	0.008	0.102	0.570	0.011
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions excluding subjects who correctly guessed that the purpose of the experiment is related to at least one of the following: (1) immigrants, (2) racism, (3) Hate. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variables in each panel from left to right are: (1) a standardized index measuring support for the shooter, (2) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (3) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (4) a dummy that equals 1 if the subject clicked on the provided links about the hate group. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.4.4.2 Inattention

One concern with online survey experiment is subjects might not pay enough attention to the survey because there is no experimenter to monitor the progress. In the worst case, subjects can simply click through the whole survey by giving random responses to every question. A typical reason for poor data quality is subjects feel they are underpaid. As discussed in Section 1.4.1, The hourly wage in my experiment is around $14, which is well above the average payment for online labor markets. In addition, subjects understand that they must answer some questions correctly to earn the highest amount of bonus payment. Finally, I include two attention checks in the survey. Subjects understand that they may not get paid if they failed the attention checks.

As shown in Section 1.4.1, 86% of the subjects passed both attention checks. Appendix Table A.15 shows that the likelihood of subjects passing attention checks are largely the same across different treatment conditions, with the exception that Republican subjects in the Hate-Background Treatment are about 9 percentage points more likely to pass the attention checks. To investigate whether the 14% of subjects who failed both attention checks biased my results, I exclude those subjects and ran the same regression specifications. Table 1.9 reports the estimates. Although the size of the magnitude is smaller, the coefficients remain statistically significant. Thus, my findings are unlikely to be driven by subjects not paying attention.

Table 1.9: Treatment Effects Excluding Subjects who Failed All Attention Checks

	A: Full Sample				B: Democrat				C: Republican			
	Index support (1)	Donation anti-immigrant (2)	Links requested (3)	Links clicked (4)	Index support (5)	Donation anti-immigrant (6)	Links requested (7)	Links clicked (8)	Index support (9)	Donation anti-immigrant (10)	Links requested (11)	Links clicked (12)
No Hate (T1)	0.187***	0.006	0.007	-0.006	0.180***	0.014	0.009	-0.003	0.214***	-0.006	0.009	-0.010
	(0.052)	(0.029)	(0.018)	(0.006)	(0.068)	(0.034)	(0.022)	(0.007)	(0.077)	(0.046)	(0.029)	(0.010)
Hate Ideology (T3)	0.121**	-0.007	0.001	0.004	0.043	-0.004	0.008	0.001	0.191**	-0.014	-0.008	0.008
	(0.053)	(0.028)	(0.018)	(0.006)	(0.069)	(0.034)	(0.023)	(0.007)	(0.077)	(0.046)	(0.029)	(0.010)
Hate Background (T4)	0.132**	-0.007	0.007	-0.008	0.243***	0.074**	0.043*	0.000	0.045	-0.087*	-0.025	-0.018*
	(0.052)	(0.028)	(0.018)	(0.006)	(0.068)	(0.034)	(0.022)	(0.007)	(0.076)	(0.045)	(0.029)	(0.010)
Control Mean	0.001	0.188	0.093	0.012	0.000	0.116	0.063	0.007	0.002	0.263	0.126	0.017
Observations	2,073	1,449	2,073	2,073	1,089	766	1,089	1,089	984	683	984	984
R-squared	0.056	0.050	0.040	0.008	0.120	0.100	0.062	0.010	0.087	0.037	0.036	0.012
T1=T3 p-value	0.204	0.635	0.718	0.090	0.045	0.601	0.962	0.571	0.765	0.876	0.567	0.082
T1=T4 p-value	0.286	0.625	0.990	0.728	0.352	0.074	0.131	0.617	0.026	0.077	0.231	0.396
T3=T4 p-value	0.828	0.996	0.706	0.040	0.004	0.021	0.124	0.945	0.053	0.103	0.537	0.009
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions excluding subjects who failed all attention checks in the survey. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variables in each panel from left to right are: (1) a standardized index measuring support for the shooter, (2) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (3) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (4) a dummy that equals 1 if the subject clicked on the provided links about the hate group. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

1.4.4.3 Recall Bias

Since the shooting I use in my experiment happened in 2019 and received nation wide attention, it is possible that subjects have seen news stories about it before. If that is the case, my results might be biased because subjects have been "treated" before. To minimize this possibility, I omitted the dates, locations, and names in the news story, except in the Hate Background Treatment, where the shooter's name is mentioned. Nevertheless, about 22% of the subjects acknowledged in the exit survey that they have heard of the shooting before. Appendix Table A.15 shows that subjects are least likely to recognize the shooting in the No Hate Treatment where the least information is given. Reassuringly, even in the Hate Ideology Treatment and the Hate Background Treatment where more information about the shooter is given, subjects are not more likely to recognize the shooting.

To investigate whether the 22% of the subjects who recognized the shooting causes bias in my results, I exclude these subjects and run the same regression specifications. The estimates are shown in Table 1.10. While the effects of the Hate Background Treatment on Democrat subjects' support for the shooter and the shooter's ideology are still evident, the treatment effects on the Republican sample are largely lost. Overall, the estimates are smaller in magnitude compared to the estimates presented in Section 1.4.2. There are two reasons that might explain why the treatment effects are weaker when excluding subjects who know the shooting before. First, since I am dropping about 22% of my whole sample, I have less statistical power to detect meaningful variations. Second, as shown in Section 1.2.2, the news coverage on hate-motivated mass shootings is intense and usually focuses on the shooter. Thus, subjects who recognized the shooting in my experiment have been treated before participating in my experiment. If there is a positive correlation between treatment intensity and treatment effect, then it explains why the treatment effects excluding subjects who have been treated repeatedly are smaller in size. This also suggests that that the treatment effects I found are likely to be an underestimate since in real life, people are likely to receive more than a one-time exposure.

Table 1.10: Treatment Effects Excluding Subjects who Recognized the Shooting

	A: Full Sample				B: Democrat				C: Republican			
	Index support (1)	Donation anti-immigrant (2)	Links requested (3)	Links clicked (4)	Index support (5)	Donation anti-immigrant (6)	Links requested (7)	Links clicked (8)	Index support (9)	Donation anti-immigrant (10)	Links requested (11)	Links clicked (12)
No Hate (T1)	0.106**	-0.014	0.002	-0.007	0.074	0.013	0.009	-0.000	0.147**	-0.042	0.002	-0.014
	(0.049)	(0.031)	(0.020)	(0.006)	(0.069)	(0.039)	(0.026)	(0.008)	(0.067)	(0.047)	(0.029)	(0.009)
Hate Ideology (T3)	0.020	-0.044	-0.008	-0.000	-0.034	0.006	0.015	0.004	0.058	-0.096**	-0.026	-0.004
	(0.050)	(0.031)	(0.020)	(0.006)	(0.070)	(0.039)	(0.027)	(0.008)	(0.069)	(0.046)	(0.030)	(0.009)
Hate Background (T4)	0.097*	0.001	0.020	-0.004	0.236***	0.086**	0.045	0.006	0.007	-0.075	0.005	-0.013
	(0.051)	(0.031)	(0.020)	(0.006)	(0.072)	(0.040)	(0.027)	(0.008)	(0.069)	(0.046)	(0.030)	(0.009)
Control Mean	0.000	0.200	0.098	0.011	-0.001	0.111	0.069	0.005	0.001	0.280	0.126	0.017
Observations	1,800	1,245	1,800	1,800	879	610	879	879	921	635	921	921
R-squared	0.063	0.042	0.035	0.005	0.144	0.090	0.053	0.011	0.098	0.038	0.038	0.013
T1=T3 p-value	0.079	0.326	0.622	0.239	0.116	0.861	0.807	0.586	0.189	0.248	0.333	0.280
T1=T4 p-value	0.867	0.622	0.357	0.608	0.021	0.060	0.181	0.444	0.040	0.473	0.915	0.928
T3=T4 p-value	0.125	0.141	0.166	0.524	0.000	0.041	0.281	0.818	0.462	0.648	0.294	0.335
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions excluding subjects who correctly guessed that the purpose of the experiment is related to at least one of the following: (1) immigrants, (2) racism, (3) Hate. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variables in each panel from left to right are: (1) a standardized index measuring support for the shooter, (2) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (3) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (4) a dummy that equals 1 if the subject clicked on the provided links about the hate group. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.4.4.4 Social Desirability Bias

One might be concerned that my results are affected by social desirability bias given that my outcomes are measuring socially inappropriate behaviors. For example, One of my main outcome variables is the donation to an anti-immigrant organization, which in many people's eyes might be socially inappropriate. Thus, people might feel prone to act in a socially desirable way, i.e., not donating to the anti-immigrant organization, which could bias my results.

To assess this possibility, I use the Marlowe-Crowne social desirability scale to measure a subject's propensity for acting in a socially desirable way (Crowne & Marlowe, 1960). This module is included as part of the demographics questions in a subset of the CloudResearch sample. I use the 7-item X1 scale as described in Fisher (1993) and construct a social desirability index using Anderson (2008)'s method.[52] Appendix Table A.16 reports the correlation between the main outcome variables and the social desirability index. Results show that subjects' social desirability index is negatively correlated with the support for the shooter. In comparison, the social desirability index does not predict the likelihood of donating to the anti-immigrant organization or the interest in hate groups, although the sign of all coefficients is negative. An additional concern is subjects' social desirability bias might be stronger in some treatments. For example, this effect could be more salient in the Hate Ideology Treatment and the Hate Background Treatment, since the news stories in both conditions hint that the shooter is motivated by anti-immigrant. If that is the case, then the estimated treatment effects should be biased downward. Following Dhar et al. (2022), I estimate the following equation.

$$\begin{aligned}Y_i = \alpha &+ \beta_1 T_{1i} + \beta_2 T_{3i} + \beta_3 T_{4i} \\ &+ \beta_4 T_{1i} * HighSD_i + \beta_5 T_{3i} * HighSD_i + \beta_6 T_{4i} * HighSD_i \\ &+ \gamma HighSD_i + \delta X_i + \epsilon_i\end{aligned} \tag{1.4}$$

Where Y_i is my outcome variable, $HighSD_i$ is an indicator variable that equals 1 if subject i has

[52] Survey script is provided in Appendix A.2.4.2.

an above-median social desirability index, and other variables are define similarly as before. The coefficients on the interaction terms between $HighSD_i$ and the treatment dummies should tell me whether there is stronger social desirability bias in different treatment conditions. Table 1.11 shows the results. Having a above-median social desirability index makes the subject significantly less likely to request links for the hate group website. This effect is not significantly for other primary outcome measures, although all signs are negative. More importantly, this effect is largely the same across different treatment conditions. The coefficients on the interaction terms are not significantly different from zero, with the exception of the No Hate Treatment. This pattern is reassuring and shows that the treatment effects are similar in magnitude for subjects with a low versus high propensity to give the socially desirable response. Thus, my results are unlikely to be affected by social desirability bias.

1.4.4.5 Multiple hypothesis testing

I have six primary outcome measures, including: 1) information demand for the shooter's manifesto, 2) information demand for the shooter's background, 3) index of support for the shooter, 4) support for the shooter's ideology measured by donation to an anti-immigrant organization, 5) whether the subject requested links to a hate group's website, 6) whether the subject clicked on the links to the hate group's website.

A common issue of having multiple outcome measures is that more hypotheses are being tested, thereby increasing the probability of false rejections. To address this concern, I compute the sharpened False Discovery Rate (FDR) q-values and report in Table 1.12 (Benjamini et al., 2006).[53] The treatment effects on support for the shooter are robust to multiple hypothesis corrections. The coefficients on Hate Ideology and Hate Background remain significant. However, the treatment effect on Democrat subjects' support for the shooter's ideology loses statistical significance. The corrected p-value is 0.164.

[53]The FDR is the expected proportion of rejections that are type I errors (false rejections). I used Anderson (2008)'s code.

Table 1.11: Robustness Check for Social Desirability Bias

	Index support		Donation anti-immigrant		Hate links requested		Hate links clicked	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No Hate (T1)	0.182	0.229	-0.124	-0.091	-0.031	-0.033	0.019	0.019
	(0.189)	(0.188)	(0.099)	(0.099)	(0.065)	(0.065)	(0.015)	(0.015)
Hate Ideology (T3)	0.217	0.260	0.056	0.071	-0.056	-0.051	0.019	0.021
	(0.188)	(0.184)	(0.097)	(0.096)	(0.065)	(0.064)	(0.014)	(0.014)
Hate Background (T4)	0.075	0.091	-0.110	-0.088	-0.075	-0.070	-0.000	-0.003
	(0.188)	(0.186)	(0.096)	(0.097)	(0.065)	(0.064)	(0.014)	(0.014)
High SD	-0.163	-0.154	-0.032	-0.031	-0.117*	-0.130**	-0.000	-0.003
	(0.186)	(0.183)	(0.094)	(0.093)	(0.064)	(0.063)	(0.014)	(0.014)
T1 * High SD	0.012	-0.077	0.248*	0.236*	-0.014	-0.023	-0.019	-0.019
	(0.263)	(0.259)	(0.134)	(0.132)	(0.090)	(0.090)	(0.020)	(0.020)
T3 * High SD	0.020	-0.069	-0.132	-0.155	0.035	0.018	-0.019	-0.021
	(0.263)	(0.258)	(0.134)	(0.132)	(0.091)	(0.089)	(0.020)	(0.020)
T4 * High SD	0.114	0.124	-0.016	-0.026	0.052	0.058	0.000	0.003
	(0.262)	(0.258)	(0.131)	(0.130)	(0.090)	(0.089)	(0.020)	(0.020)
Constant	0.100	-0.183	0.241***	-0.173	0.205***	0.033	0.000	-0.057***
	(0.139)	(0.278)	(0.073)	(0.143)	(0.048)	(0.096)	(0.011)	(0.022)
Observations	399	399	280	280	399	399	399	399
R-squared	0.015	0.073	0.046	0.104	0.027	0.073	0.014	0.048
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table presents results from OLS regressions. SD Index is an index ranging from 0 to 1 that measures a subject's propensity to give the socially desirable response. The dependent variables in each panel from left to right are: (1) a standardized index measuring support for the shooter, (2) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (3) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (4) a dummy that equals 1 if the subject clicked on the provided links about the hate group. High SD is a dummy that equals 1 if the subject has a above-median social desirability index. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, a dummy for the Hate Background Treatment, High SD, and the interaction terms between the treatment dummies and High SD. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

1.4.5 Discussion of Effect Sizes

In previous sections, I showed that subjects from the Democrat sample significantly increase their support for the shooter and the shooter's anti-immigrant ideology when they read a news story that contains information about the shooter's background. It is worth pointing out that both the effect size and the baseline level are relatively small. For example, when asked if they admire the shooter's courage on a 5-point Likert scale, about 91 percent of all the subjects chose strongly

Table 1.12: Correction for Multiple Hypothesis Testing

	(1) Demand manifesto	(2) Demand background	(3) Index support	(4) Donation anti-immigrant	(5) Hate links requested	(6) Hate links clicked
		Panel A: Full sample				
No Hate	0.050	0.055	0.194	-0.008	0.009	-0.007
	(0.078)	(0.052)	(0.000)	(0.773)	(0.611)	(0.242)
	[0.145]	[0.108]	[0.001]	[0.660]	[0.619]	[0.300]
Hate Ideology	-0.087	-0.058	0.115	-0.017	0.005	0.007
	(0.002)	(0.044)	(0.017)	(0.524)	(0.762)	(0.254)
	[0.012]	[0.107]	[0.050]	[0.577]	[0.660]	[0.300]
Hate Background	-0.036	-0.125	0.126	-0.016	0.012	-0.005
	(0.199)	(0.000)	(0.009)	(0.549)	(0.489)	(0.409)
	[0.284]	[0.001]	[0.035]	[0.577]	[0.577]	[0.518]
Control Mean	0.436	0.557	0.001	0.198	0.098	0.012
Observations	2,400	2,400	2,400	1,665	2,400	2,400
R-squared	0.031	0.024	0.072	0.045	0.040	0.009
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
		Panel B: Democrat sample				
No Hate	0.047	0.056	0.169	0.021	0.012	-0.003
	(0.235)	(0.167)	(0.011)	(0.526)	(0.585)	(0.719)
	[0.778]	[0.57]	[0.064]	[1]	[1]	[1]
Hate Ideology	-0.026	-0.004	0.034	0.013	0.013	0.004
	(0.511)	(0.927)	(0.609)	(0.697)	(0.562)	(0.587)
	[1]	[1]	[1]	[1]	[1]	[1]
Hate Background	-0.030	-0.130	0.264	0.069	0.039	0.004
	(0.448)	(0.001)	(0.000)	(0.037)	(0.078)	(0.600)
	[1]	[0.011]	[0.002]	[0.164]	[0.281]	[1]
Control Mean	0.407	0.563	0.000	0.114	0.067	0.007
Observations	1,199	1,199	1,199	842	1,199	1,199
R-squared	0.032	0.033	0.118	0.087	0.057	0.013
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
		Panel C: Republican sample				
No Hate	0.057	0.060	0.234	-0.037	0.012	-0.011
	(0.155)	(0.140)	(0.001)	(0.377)	(0.660)	(0.254)
	[0.275]	[0.275]	[0.005]	[0.477]	[0.591]	[0.378]
Hate Ideology	-0.146	-0.111	0.187	-0.045	0.002	0.010
	(0.000)	(0.006)	(0.005)	(0.274)	(0.952)	(0.264)
	[0.005]	[0.021]	[0.021]	[0.378]	[0.699]	[0.378]
Hate Background	-0.038	-0.114	0.043	-0.093	-0.007	-0.013
	(0.350)	(0.005)	(0.529)	(0.022)	(0.790)	(0.162)
	[0.477]	[0.021]	[0.478]	[0.052]	[0.699]	[0.275]
Control Mean	0.465	0.551	0.001	0.278	0.130	0.017
Observations	1,201	1,201	1,201	823	1,201	1,201
R-squared	0.039	0.027	0.113	0.034	0.036	0.013
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions. All variables follow the same notation as before. Standard p-values are reported in parentheses, under the coefficients. FDR-adjusted p-values, computed following Anderson (2008) are reported in square brackets below.

disagree (a value of 1). Similarly, when asked if they think the shooter's action can be justified, about 84 percent of the subjects chose 1. Despite the small magnitude, my results are meaningful for two reasons.

First, as discussed in the last section, the treatment effects detected from my experiment are likely to be underestimated since it's a one time exposure. In real life, people will likely encounter news stories about the shooter's ideology or background repeatedly. Thus, the small magnitude of treatment effects does not diminish its significance. Admittedly, the index of support for the shooter is constructed using self-reported attitudinal responses, and thus does not necessarily predict change in behaviors. However, literature in criminology shows mass shooters may find inspiration from previous shooters. This connection often starts from admiration and sympathy, then becomes stronger over time Langman (2018). Thus, it is possible that the self-reported increase in support for the shooter could translate into behavioral change in the future. To investigate this possibility, Appendix Table A.17 shows support for the shooter strongly predicts support for the shooter's ideology and interest in hate groups. Therefore, even though the observed treatment effects might be driven by a small fraction of subjects, it is nevertheless alarming.

Second, it should be noted that the observed treatment effect on donation rate is comparable with other papers using similar methodology. The review paper by Haaland et al. (2020) points out that the effect size on behavioral measures are typically small in information experiments. While donation to a charity organization is commonly used as an outcome measure in information experiments, many studies find very little difference between the treatment group and the control group. For example, Settele (2022) uses donation to an NGO for women right, Grigorieff et al. (2020) uses donation to a pro-immigrant charity, Haaland and Roth (2021) uses donation to a pro-black civil rights organization. However, all three studies do not find statistically significant treatment effect on donation. Two exceptions are Bursztyn, Egorov, and Fiorin (2020), who found an increase of 50 percent in donation rate to an anti-immigrant organization, and Bursztyn, Haaland, et al. (2020), who found an increase of 23 percent in donation rate to a pro-black organization. In comparison, I find a 62 percent increase in donation rate to an anti-immigrant organization from the Democrat sample. This magnitude is larger than Bursztyn, Egorov, and Fiorin (2020) and Bursztyn, Haaland, et al. (2020).

1.5 Conclusion

In this paper, I study whether news coverage of hate-motivated mass shootings increases hatred. In the first part of my analysis, I use observational data from multiple sources to provide evidence that 1) hate-motivated mass shootings receive higher media coverage which often focuses on the shooter; 2) people show higher interest in hate-motivated mass shootings, and in particular, the shooter; 3) immediately following a hate-motivated mass shooting targeting a specific group, there's an increase in the number of hate crimes against the same victimized group. Based on these findings and guided by the existing literature, I hypothesize that the way the media covers hate-motivated mass shootings causally generates more hatred. In the second part of my analysis, I employ an online experiment to test my research hypothesis. In the experiment, subjects are asked to read a piece of news story about the 2019 El Paso shooting that targeted Hispanic immigrants. Each subject is randomly assigned to one of the four treatment conditions that vary in the level of informativeness, i.e., whether the news story discloses the shooting was targeting Hispanic immigrants, and whether it covers the shooter's hateful ideology (white supremacy, anti-immigrant) or background (name, photo, childhood). I then measure subjects' interest in the shooter, attitudes toward the shooter, the shooter's ideology, and interest in a white supremacy hate group.

My first finding from the experiment is that subjects are not more interested in hate-motivated than non-hate-motivated shootings. This suggests that the higher public interest in hate shootings I observed in the search data is likely due to the fact that these crimes receive more media coverage, rather than to the fact that subjects are intrinsically more interested in them. My second finding is that providing more information on the shooter's background significantly increases support for the shooter from Democrat subjects, providing more information on the shooter's ideology significantly increases support for the shooter from Republican subjects. My third finding is that consistent with the increase in support for the shooter, Democrat subjects who read the news story that emphasizes the shooter's background significantly increase their support for the shooter's anti-immigrant ideology as measured by donations to anti-immigrant organization. My fourth finding is there is suggestive evidence that the news story with emphasis on the shooter's background increases Democrat subjects interest in a white supremacy hate group.

Overall, this paper shows that news coverage of mass shootings could have unintended consequences. In particular, news stories could positively affect viewers' attitude toward the shooter, and negatively affect viewers' attitude toward the victims. Thus, my findings provide support for the argument that media coverage of sensitive topics should be regulated. This paper has implications for future work. To start with, my experiment studies the reaction to a specific mass shooting from a specific group of subjects. Subsequent research should examine the reaction to a different mass shooting from a broader audience. Second, my experiment focuses on how different types of media coverage change viewer's attitudes toward the victimized group in the shooting. It would also be interesting to see whether media coverage changes how the victimized group feels toward the shooter's group. Again, consider the 2019 El Paso shooting as an example, will immigrants become more resentful toward white people after they saw news stories about the shooter's white supremacy ideology? Third, my treatments vary in the amount of information about the shooter. Many people argue that media should shift attention from shooters to victims and survivors.[54] Future work should investigate whether and how news coverage that emphasizes the victims' stories and backgrounds may affect viewers' attitudes toward the victimized group and the shooter's ideology, possibly leading to less hatred.

[54]For example, see the No Notoriety Campaign, and `https://www.reportingonmassshootings.org/`

2. CAN SOCIAL MEDIA RHETORIC INCITE HATE INCIDENTS? EVIDENCE FROM TRUMPS "CHINESE VIRUS" TWEETS

2.1 Introduction

[1] Just how far-reaching is the influence of high-profile individuals and what sorts of behaviors can they alter? Research has shown that high-profile individuals can affect consequential pro-social behaviors like interest in preventative health care (Cram et al., 2003; Evans et al., 2014; Roberts & Dusetzina, 2017; Alatas et al., 2019) and voting (Jackson & Darrow, 2005; Austin et al., 2008; Garthwaite & Moore, 2013; Chou, 2015; Xiong, 2021). In this paper, we investigate whether this sort of influence can extend to anti-social behaviors as well. The answer to this question is increasingly relevant given ongoing debates about restrictions on the freedom of speech in instances in which that freedom may cause harm. Perhaps most visible among these recent debates is President Donald Trump's use of social media prior to the storming of the United States Capitol, followed by his subsequent suspensions from Twitter due to concerns about "further incitement of violence" [2] and from Facebook,

Instagram, Snapchat, and Twitch due to similar concerns.[3] Facebook has subsequently changed its policies to allow less leniency for public figures and will consider reinstating his accounts in January 2023 "when it will look to experts to decide whether the risk to public safety has receded"; in the meantime, Trump has asked federal courts to require Twitter to reinstate his account on the grounds of unfair censorship.[4] While it is infeasible to disentangle the contribution of Trump's speech from other factors that may have contributed to the Capitol Hill violence, the president's

[1]This chapter is a joint work with Jason M. Lindo (jlindo@tamu.edu), And Jiee Zhong (jieezhong@tamu.edu) from Texas A&M University. Corresponding author: Zhong. We gratefully acknowledge Thomas Fujiwara, Karsten Müeller, and Carlo Schwarz for sharing their data on the number of Twitter users across counties, the Stop AAPI Hate reporting center for making available their data on anti-Asian incidents, and Bing He, Caleb Ziems, Sandeep Soni, Naren Ramakrishnan, Diyi Yang, and Srijan Kumar for making available their code and data. We also thank Andrew Barr and Daniel Sturm for their detailed comments.

[2]See here for media coverage.

[3]See here for media coverage. More recently, Twitter permanently suspended Marjorie Taylor Greene's personal account in January 2022 for repeated violations of its COVID-19 misinformation policy.

[4]See here for media covearge.

earlier use of Twitter during less volatile and consequential periods provides an opportunity to understand whether the speech of high-profile individuals may incite anti-social behavior more generally. Towards this end, in this study we consider whether President Donald Trump's remarks about China on Twitter during the COVID-19 pandemic led to an increase in the number of anti-Asian incidents in the subsequent days.

Our analysis focuses on incidents that occurred around the time Trump began attributing COVID-19 to China. We use data on incidents from the Stop AAPI Hate reporting center, which tracks incidents of hate, violence, harassment, discrimination, shunning, and child bullying against Asian Americans and Pacific Islanders in the United States.[5] Our analysis of national trends shows an extremely large spike in incidents on March 20, 2020. We argue that this spike is indicative of a causal effect of Trump's influence given its timing relative to Trump's initial references to the "Chinese Virus" (one tweet on March 16, another on March 17, followed by four on March 18), which were followed by a spike in the number of anti-Asian Covid-19 tweets on March 19 (He et al., 2021). Further supporting this interpretation of the results, we show that Google search queries for "Chinese Virus" also spiked on March 19, the day before the spike in anti-Asian incidents; that "Trump" and "Trump Chinese Virus" are the search queries most closely related to search queries for "Chinese Virus." Moreover, in difference-in-differences and event-study analyses leveraging spatial variation, we find that the spikes in anti-Asian Covid-19 tweets and anti-Asian incidents are more pronounced in counties that supported Donald Trump in the 2016 presidential election relative to those that that supported Hillary Clinton.

Our work complements a handful of other studies that have examined anti-social effects of the specific content disseminated through media, including research showing that radio programming in Rwanda calling for the extermination of the Tutsi minority had a significant impact on participation in killings by militia groups and ordinary civilians (Yanagizawa-Drott, 2014); that radio content incited anti-Semitic and pro-Nazi acts by ordinary citizens (Adena et al., 2015; Wang, 2021); and that the fictional portrayal of the KKK in the film The Birth of a Nation caused lynch-

[5]A report that provides victim narratives describing these types of incidents for incidents reported through March 25, 2020 can be found here.

ings and race riots in the United States in the early 20th century (Ang, 2020). Our work also complements recent research on social media showing that county-level Twitter penetration reduced the Republican vote share in the 2016 and 2020 presidential elections without having any effects on Congressional elections and previous presidential elections (Fujiwara et al., 2021).

The remainder of this paper is organized as follows. In the next section, we describe the data on reports of anti-Asian incidents. We then discuss the context surrounding Trump's "Chinese Virus" tweets and national trends in Section 2.3. Section 2.4 presents the results of our analyses comparing Trump-supporting counties and Clinton supporting counties. In subsequent sections, we consider mechanisms, discuss the limitations of our analyses, and conclude.

2.2 Reports of Anti-Asian Incidents

Since the Coronavirus outbreak in 2020, the number of reported anti-Asian hate crime incidents has risen dramatically among most major cities in the United States, including New York City which saw an 833% increase in racially motivated crimes against Asian Americans (Levin, 2021). This is of particular concern because Asian Americans are one of the most highly urbanized segments of the U.S. populations, with approximately 95 percent living in urban areas.[6]

Our analysis of Anti-Asian incidents is based on self-reports of incidents to the "Stop AAPI Hate" Reporting Center, from the beginning of 2020 through April of 2020. The Stop AAPI Hate webpage (stopaapihate.org) has a very simple layout that allows visitors to immediately begin reporting an incident (Appendix Figure B.1). Those reporting incidents are asked a total of 18 questions including: the date the incident occurred, the kind of incident they experienced (based on 10 categories), and their state and zip code, in addition to other details about their experience, demographics, and contact information.

In terms of the subcategories comprising these incidents, 54 percent involved verbal harassment or name-calling, 18 percent involved avoidance/shunning (e.g., deliberate avoidance of, distancing from, or social rejection for racial/ethnic group), 12 percent involved physical assault (including being coughed or spat on), 4 percent involved a workplace discrimination, 3 percent involved

[6]See this report from the Population Reference Bureau.

refusal of service (at a business establishment, public transit, or private transportation such as ride-share services), and 2 percent involved online misconduct, with the remaining 8 percent of incidents in other categories.[7] Moreover, nearly all the incidents in our data were reported by victims who directly experienced an incident in person. This is a key distinction from studies that have considered online hate speech, such as He et al. (2021), which is usually not directed at any specific individual and may not be seen by members of the disparaged group.

Though we expect these data to substantially understate the degree to which these incidents occur across the United States, we view them as providing a useful proxy for such incidents. We discuss the limitations of these data—and the implications for the interpretation of the results of our analyses—in Section 2.6.

2.3 Context and National Trends

Figure 2.1 shows the number of incidents reported as occurring on a given date (not the date of the report) from January 2020 through May 2020. For context, Figure 2.1 also plots a measure of (US-based) web search activity for "Chinese Virus" based on Google Trends[8] and Table 2.1 provides a timeline of significant events related to the pandemic and the use of potentially stigmatizing language. We discuss the patterns in these data in the following subsections. Note that similar patterns are evident if daily incident counts are adjusted so that they are relative to the average number observed on the same day of the week over the analysis window (Appendix Figure B.3). They are also similarly evident in urban areas and rural areas (Appendix Figure B.4).

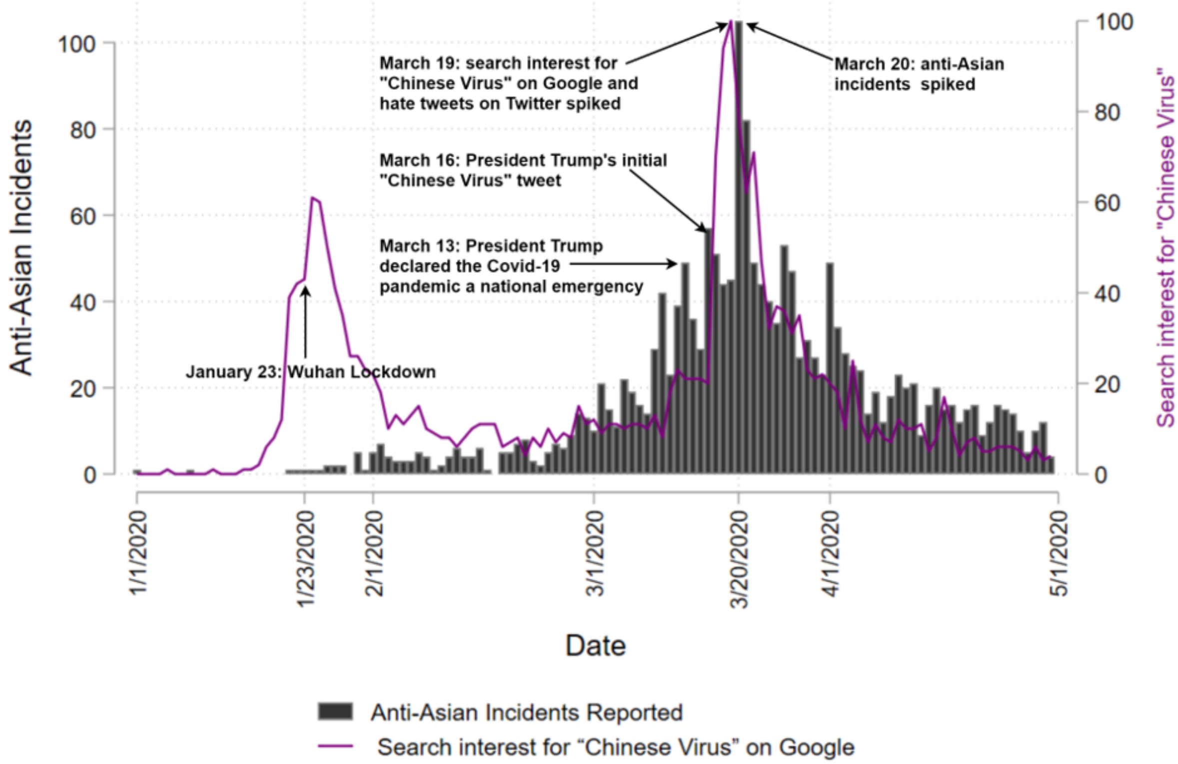
2.3.1 Before Trump's Initial "Chinese Virus" Tweet

The number of incidents was fairly stable from late January throughout most of February and then began to rise rapidly towards the end of February and into March. Over this period of time, concerns about the pandemic were escalating as the virus spread throughout China and then to other parts of the world. On January 23, Chinese authorities implemented a lockdown for the

[7] These category descriptions are shown to individuals reporting incidents. See Appendix Figure B.1

[8] Google Trends provides access to search requests made to Google. The data is aggregated and normalized. Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics.

Figure 2.1: Anti-Asian Incidents and Google Search Interest in “Chinese Virus” Over Time



Notes: This figure plots the trend in anti-Asian incidents and the popularity of the term “Chinese Virus” against a timeline. The bars in gray show the number of hate incidents reported as having occurred on a given day. Hate incidents include both hate crimes as well as hateful acts that are not legally defined as crimes, such as verbal harassment, shunning, and refusal of service at restaurants. The purple line chart shows the interest over time for the term “Chinese Virus” on Google. Numbers represent search interest relative to the highest point (100) on the chart for the given time in the United States. Data are from “Stop AAPI Hate” and Google Trends.

city of Wuhan. One week later, the World Health Organization declared the virus an international emergency. Days later, the United States declared a public health emergency. These data show a growing number of incidents over time, a dramatic spike on March 20, and a subsequent decline.

Google Trends shows a rise and fall in search queries for “Chinese Virus” around the time of the Wuhan lockdown. This fact highlights that this term was in use well before Trump used the term in public. Recognizing the potential for this sort of language to do harm, on February 11 the World Health Organization recommended the use of “coronavirus” and “COVID-19” to describe the virus instead of potentially stigmatizing alternatives (World Health Organization 2020). Having already

Table 2.1: Timeline of Events

Date	Event
1/23/2020	Chinese authorities lockdown Wuhan
1/30/2020	WHO declares coronavirus an international emergency
2/3/2020	U.S. declares public health emergency
2/11/2020	WHO issues guidance to use “coronavirus” and “COVID-19,” and to avoid “stigmatizing”
3/1/2020	The first case of COVID-19 in New York during the pandemic is confirmed
3/6/2020	US Secretary of State Pompeo uses “Wuhan virus” on Fox and Friends and CNBC
3/8/2020	US Congressman Paul Gosar uses “Wuhan virus” on Twitter
3/9/2020	US Congressman Kevin McCarthy uses “Chinese coronavirus” on Twitter
3/11/2020	WHO declares COVID-19 a pandemic
3/13/2020	President Trump declares the Covid-19 pandemic a national emergency
3/16/2020	Trump’s initial “Chinese Virus” Tweet
3/17/2020	Trump’s second “Chinese Virus” Tweet
3/18/2020	Trump has four “Chinese Virus” Tweets on this single day
3/19/2020	Trump uses “Chinese Virus” in a press conference and responds “it’s not racist at all” when asked
3/19/2020	Hate tweets spike on Twitter; Google search queries for Chinese virus spike
3/20/2020	Anti-Asian incidents spike

Sources: The American Journal of Managed Care Staff (2021), Darling-Hammond et al., (2020), He et al. (2021).

fallen from its earlier levels around the time of the Wuhan lockdown, search interest for “Chinese Virus” remained steady throughout most of February and the first half of March. This is particularly notable in light of the fact the first case of COVID-19 in the United States was confirmed on March 1 and some U.S. political officials used the terms “Chinese Virus” and “Wuhan Virus” in public during this period of time.[9]

2.3.2 Trump’s Initial “Chinese Virus” Tweets and the Immediate Aftermath

On March 13, Donald Trump declared a national emergency concerning the COVID-19 pandemic. On March 16, he used the term “Chinese Virus” for the first time in public, in a tweet about his intent to support industries affected by the pandemic.[10] A day later he used the term again, this time in the context of highlighting that the effects of the pandemic varied across states (in an appar-

[9]Specifically, on March 6, the former Secretary of State Mike Pompeo was interviewed on Fox and Friends, a morning news show hosted by Fox News Channel. During the interview, Pompeo repeatedly addressed the coronavirus as the “Wuhan Virus” in addition to criticizing the Chinese government for lack of transparency and false information. This marked the first time that anyone from the Trump administration used such language in public. Congressman Paul Gosar also used the term “Wuhan Virus” on Twitter on March 8 and congressman Kevin McCarthy used the term “Chinese coronavirus” on Twitter on March 9.

[10]Data for Trump’s tweets is collected from Trump Twitter Archive. This archive checked Twitter every 60 seconds to record every Trump tweet into a database.

ent effort to argue for state-specific responses rather than a federal response). One day later, there was a marked increase in the intensity with which Trump used potentially inflammatory language. In particular, he used the term "Chinese Virus" in four separate tweets on March 18. Moreover, in this set of tweets Trump: referenced the "onslaught of the Chinese Virus"; stated that it was "not your fault!" to people who were out of work; stated that he did "a very good job. . . to close the 'borders' from China"; and explained that he "signed the Defense Production Act to combat the Chinese Virus." The full text of these tweets is shown in Figure 2.2.

Figure 2.2: Trump Initial Tweets Referencing the "Chinese Virus"

These tweets would have been seen by a very large number of Americans.[11] Surveys before the pandemic indicate that 19 percent of adult Twitter users in the United States "followed" Trump on Twitter, including 31 percent of Republicans and 13 percent of Democrats (Wojcik et al., 2019). Naturally, many more people are exposed to Trump's tweets via retweets, quote tweets, media coverage, and personal interactions. Relative to Trump's other original tweets during this period of time, Trump's first tweets referring to the "Chinese Virus" were popular. Trump posted 35 original tweets from March 16 to March 18 and these tweets averaged 146,737 likes and 29,196 retweets.

[11]Notably, over this period of time U.S. media coverage of COVID-19 arguably focused disproportionately on Trump, and was more negative in tone compared to non-U.S. media (Sacerdote et al., 2020).

His tweets referring to the “Chinese Virus” exceeded these numbers with an average of 185,956 likes and 38,907 retweets.[12]

The next day, on March 19, Trump used the term on television for the first time—in a coronavirus taskforce press conference—and declared “it’s not racist at all” when confronted by a reporter. This instance led to especially heightened attention because photos from the press conference showed that his notes had “corona” and “coronavirus” crossed out and replaced with “Chinese”.[13]

Web-based search queries for “Chinese Virus” began to spike the day after Trump first used the term in public and reached an apex on March 19, the same day he used the term in a press conference and the day after he used the term in four separate tweets (Figure 2.1). While it is possible that this timing could be coincidental, Google Trends data on related queries (Table 2.2) suggests that this is highly unlikely. In particular, Google Trends information on “related queries” captures the degree to which users search for different terms during a search session.[14] Based on these data, searches for “Chinese Virus” were not strongly associated with Trump before he used the term in public. From the beginning of the year through March 15, the top five search queries related to “Chinese virus” were: corona virus, Chinese corona virus, coronavirus, Chinese virus 2020, and China virus. This changed after Trump’s used the term on March 16. From March 16 to March 20, the top search queries related to searches for “Chinese Virus” were: Chinese virus Trump, Trump, the Chinese virus, Donald Trump Chinese virus, and Trump Twitter. From March 21 through the end of April, searches for “Chinese virus Trump”” continued to be among those most closely related to searches for “Chinese virus,” only trailing “the Chinese virus.”

Along similar lines, data from He et al. (2021) shows that the number of hateful anti-Asian COVID-19 tweets from Twitter users in the United States began to spike the day after Trump first

[12] Trump also retweeted 65 times over these days. Including these retweets, his 100 posts on Twitter across these days averaged 51,358 likes and 14,492 retweets. All of these numbers are authors’ calculations.

[13] See here for media coverage.

[14] Google does not disclose how they define a search session. Generally, a search session consists of all the search requests from a user within a given timeframe. A session lasts until there’s inactivity. A common value for the inactivity threshold is 30 minutes and is sometimes described as the industry standard.

Table 2.2: Top Search Queries from the United States Related to Searches for "Chinese Virus"

1/1/20 - 3/15/20	3/16/20 - 3/20/20	3/21/20-4/30/20
corona virus (100)	chinese virus trump (100)	the chinese virus (100)
chinese corona virus (98)	trump (98)	chinese virus trump (83)
coronavirus (59)	the chinese virus (45)	new chinese virus (30)
chinese virus 2020 (36)	donald trump chinese virus (11)	donald trump chinese virus (2)
china virus (29)	trump twitter (6)	

Notes: Each column displays Google Trends "top search queries" related to searches for "Chinese Virus" over the specified period of time, based on different queries that occur in the same "search session." Numbers in parentheses represent the degree to which searches for the phrase are related to searches for "Chinese Virus" during the same search session, on a Google Trends 0-100 scale. This analysis focuses on US-based searches only.

used the term in public (Figure B.2).[15] Moreover, the number of such tweets reached its highest point on the same day as the searches for "Chinese Virus."[16]

As shown in Figure 2.1, the very next day (March 20) there is a clear and dramatic spike in the number of anti-Asian incidents, 125.35% percent higher than the day prior. Though the number of incidents remains elevated the following day (March 21) they subsequently return to prior levels. This pattern is consistent with the pattern of ant-Asian hate speech observed on Twitter (Figure A2). Trump faced criticism for his use of this term in subsequent days, he did not use the term in his next press conference, and on March 23 he tweeted "it is very important that we totally protect our Asian American community in the United States, and all around the world."

The same general pattern in incidents is evident for verbal harassment or name-calling (Figure B.5, Panel A) and avoidance/shunning (Figure B.5, Panel B), i.e., the two most frequent incident types. Patterns are more difficult to discern for less-frequent incident types (Figures B.6 through

[15]He et al. (2021) constructed this dataset by collecting all covid-related tweets using keywords such as "covid" and "corona" and then classifying which tweets involved anti-Asian hate. The latter was done using a classifier that was trained on a subset of hand-labeled tweets using machine learning.

[16]Along similar lines, Hswen et al. (2021) document that half of tweets with the hashtag #chinesevirus showed anti-Asian sentiment versus one fifth of tweets with the hashtag #covid19; that anti-Asian sentiment in tweets with these hashtags was greater the week after March 16 than the week before; and that this growth in anti-Asian sentiment was significantly larger for tweets using the hashtag #chinesevirus. In addition, Crisis Text Line, a group that provides free mental health support via text message, saw a 50% increase in texts received from people identifying as Asian the week of March 16, after that number of texts had remained fairly stable at around 2,200 per week during the preceding month (see here for media coverage).

B.7), such as physical assault, workplace discrimination, refusal of service, and online misconduct. That said, we note that the daily high for incidents of physical assault coincided with the day that Trump first used "Chinese Virus" publicly and the second daily high coincided with the spike in incidents overall (Figure B.6, Panel A).

2.4 Estimates Comparing Trump- vs Clinton-Supported Counties

If the spike in incidents is a result of Trump's influence, we would expect to see a larger spike in anti-Asian behavior in areas where Trump has greater support.

To investigate this, we compare counties won by Donald Trump in the 2016 presidential election relative to those won by Hilary Clinton.[17] We focus on incident rates per 100,000 Asian residents for counties with at least one incident reported following January 1, 2020.[18] In Figure 2.3, we show how this incident rate changed over time, separately for Trump-supported and Clinton-supported counties. For comparison, we also show how hateful anti-Asian COVID-19 tweets changed over time across these sets of counties.[19] For both Trump-supported counties and Clinton-supported counties, these figures show a spike in anti-Asian COVID-19 tweets following Trump's initial "Chinese virus" tweet, which was followed by a spike in anti-Asian incidents. Both spikes are larger in Trump-supported counties. County-level averages for the two weeks before and after Trump's initial "Chinese virus" (Table B.1) show a similar pattern of elevated incidents (and anti-Asian COVID-19 tweets) that is larger in Trump-supported counties than in Clinton-supported counties.

A difference-in-differences estimate based on the same county averages indicates that anti-Asian Covid-19 tweets spiked by 11.9 per 100,000 Asian residents in Trump-supported counties, over and above the increase observed in Clinton-supported counties.[20] Put in other terms, anti-

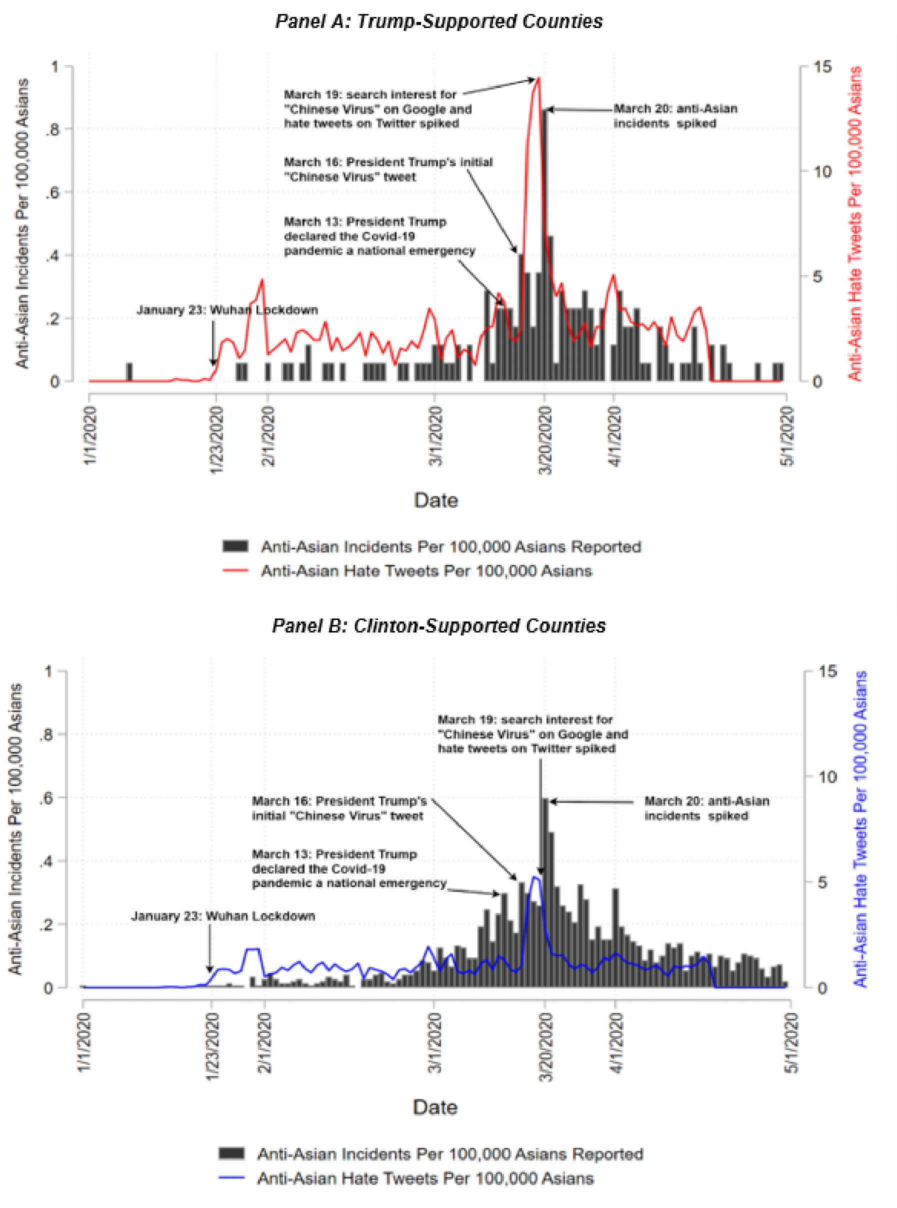
[17] Election outcomes are from the New York Times: `https://www.nytimes.com/elections/2016/results/president`

[18] Data on the number of Asian residents are 2019 estimates from the U.S. Census Bureau, data series CC-EST2019-ALLDATA-[ST-FIPS].

[19] These are from the same Twitter data described above, with counties geocoded based on the latitude and longitude of each tweet. Note that not all tweets have this information because users can turn off the GPS option from their settings. Since it is unlikely that users would do so in a manner coinciding with Trump's "Chinese virus" tweets, we do not think this is a serious issue for our purposes.

[20] This calculation reflects the day-of-spike average for Trump-supported counties minus the pre-"Chinese Virus"

Figure 2.3: Daily Anti-Asian Incidents and Anti-Asian COVID-19 Posts on Twitter, Based on County Support of Trump vs Clinton in 2016 Election



Asian tweets spiked 235 percent more in Trump-supported counties than in Clinton-supported counties. Similar calculations indicate that anti-Asian incidents spiked by 58.7 per 100,000 Asian residents in Trump-supported counties, over and above the increase observed in Clinton-supported counties. Again put in other terms, anti-Asian incidents spiked by over 4000 percent more in

tweet average for Trump-supported counties (22.389 - 2.971 = 19.418), minus the same difference for Clinton-supported counties (9.289-1.790 = 7.499).

Trump-supported counties than in Clinton-supported counties.

We also investigated this issue with an event-study specification that compares how anti-Asian behavior changes over time in Trump-supported counties relative to Clinton-supported counties, using data on incidents from 14 days before Trump's first "Chinese Virus" tweet through 30 days after that tweet. We estimate the following specification via ordinary least squares:

$$y_{ct} = \sum_{k-14, k \neq -1}^{30} \beta_k Trump2016_c * I(k = Date_t - Mar16) + \theta X_{ct} + \alpha_c + \gamma_t + \epsilon_{ct}$$

The outcome variable y_{ct} is the number of anti-Asian incidents that occurred on a particular day in a particular county per 100,000 Asian residents. We include county fixed effects α_c, which control for fixed differences across counties in the rate of anti-Asian behavior, and date fixed effects γ_t, which control for changes over time in the rate of anti-Asian behavior experienced in all counties. We also control for the logarithm of daily Covid-19 cases (cumulative) across counties and over time, which allows for within county changes in the spread of the virus to influence anti-Asian behavior directly, though the estimates are nearly identical if this control variable is omitted.[21] We are primarily interested in the coefficients β_k on the interactions between an indicator variable for whether a county supported Donald Trump in the 2016 presidential election ($Trump2016_c$) and k indicator variables $I(k = t - Mar16)$ for being k days from March 16, the date of Trump's first tweet to mention "the Chinese virus." This set of coefficients reflects the degree to which anti-Asian incidents are differentially elevated for Trump-supporting counties relative to other counties k days from Trump's tweet, over and above differences that are expected based on the differences that are observed across counties at other times, the changes observed across all counties over time, and the changes that are expected based on the fluctuating number of Covid-19 cases within counties over time.[22] Given that Trump's tweets may have also affected Clinton-supporting counties

[21] Covid data are from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins: `https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data`

[22] Following A. C. Cameron et al. (2011), the 95% confidence intervals we report are based on two-way clustered standard error estimates that allow errors to be correlated within counties over time and also across counties on the same date.

(as suggested by the results shown in Figure 2.3 and Table B.1), these coefficient estimates may understate the overall effect on anti-Asian incidents.

Figure 2.4 plots the resulting β_k estimates, with circles representing estimates corresponding to days before Trump's first "Chinese virus" tweet and triangles representing estimates corresponding to subsequent days.[23] The estimated effects are largely close to zero for the days before Trump's first "Chinese virus" tweet, and even on the day of that tweet, indicates similar pre-tweet trends in anti-Asian incidents between counties with greater and lesser support for Donald Trump. Following the president's initial tweets (one on March 16, another on March 17, and four on March 18) and the subsequent spike in anti-Asian tweets and "Chinese virus" searches (on March 19), there is a large spike in the rate of anti-Asian incidents in counties that supported Donald Trump in the 2016 presidential election relative to those that supported Hilary Clinton. Indeed, the point estimate indicates an additional 60 out of every 100,000 Asians reported being victimized in Trump-supporting counties, over and above changes in Clinton-supporting counties, which is nearly identical to the difference-in-differences estimate reported above.

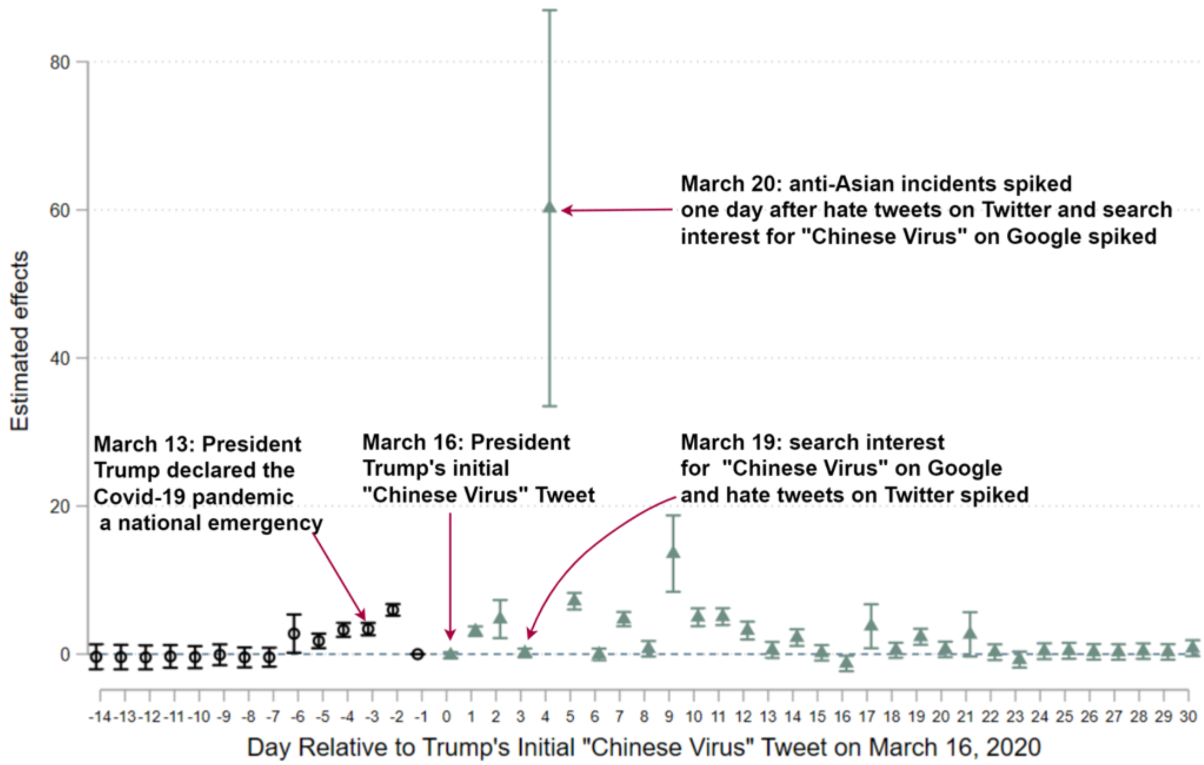
When we exclude counties in which the vote share was evenly split between Trump and the other candidates, and thus increase the contrast in support for Donald Trump across groups in our sample, we see even larger relative spikes in reported anti-Asian behavior in Trump-supporting counties.[24] An analysis by week (instead of day) illustrates that this spike was not just a shift in the timing of incidents that might have otherwise happened at a slightly later date. Specifically, there was a relative spike in anti-Asian incidents in Trump-supporting counties the week following Trump's initial tweets, it remained elevated the following week, and then fell back to parallel the trend in Clinton-supporting counties (Figures B.12 and B.13).[25]

[23] See Appendix Figure B.9 for estimates based on the specification that omits the control for Covid-19 cases.

[24] Relative to these estimates using all counties, Appendix Figure B.10 shows a spike nearly twice as large if the analysis excludes counties where the share voting for Donald Trump was 40-60 percent and a spike more than three times as large if the analysis excludes counties where the share voting for Donald Trump was 30-70 percent.

[25] We note that the weekly estimates depicted in this figure are flat and virtually identical for weeks -10 to -2 and weeks 3-30, which supports the common trends assumption that the analysis replies upon. We note that these estimates are not centered on zero, however, because incidents rose slightly more in Trump-supporting counties than Clinton-supporting counties in the week prior to Trump's first "Chinese Virus" tweet (which serves as the omitted period for the event-study specification), which was also evident in Figure 2.4.

Figure 2.4: Estimated Effects on Anti-Asian Incidents in Trump- vs Clinton-Supported Counties



Notes: This figure plots the estimated effects of Trump's initial "Chinese virus" tweet on anti-Asian incidents in counties that supported Trump in 2016 versus those that supported Clinton. Estimates control for the logarithm of the total number of Covid-19 cases plus one, county fixed effects, and date fixed effects. The outcome variable is the number of anti-Asian incidents per 100,000 Asian residents. Data, restricted to incidents fourteen days before Trump's initial "Chinese virus" tweet and 30 days after that tweet, are from the Stop AAPI Hate database. Confidence intervals are based on two-way standard-error estimates allowing for clustering within counties over time and across counties on the same date A. C. Cameron et al. (2011).

Just as nationwide trends in overall incidents were mirrored by incidents of verbal harassment or name-calling and avoidance/shunning, our analyses of Trump- and Clinton-supporting counties indicates that the general pattern for overall incidents (Figure 2.4) is mirrored in analyses of verbal harassment or name-calling and avoidance/shunning (Figure B.13, Panel A). That said, the estimated effect on shunning is an order of magnitude larger than the estimated effect on verbal assault. We also find significant effects on the very same day for refusal of service (Figure B.14, Panel A). There is little evidence of effects on physical assault, workplace discrimination, or online misconduct (Figure B.14, Panel B and Figure B.15).[26]

2.5 Mechanisms

In terms of the mechanisms underlying these effects, one may wonder if the difference-in-differences and event-study estimates may have resulted from differential exposure to Trump's "Chinese virus" tweets as opposed to differential responses conditional on exposure. We have investigated this possibility in several ways. First, we used Twitter-usage data from Fujiwara et al. (2021), generously shared by the authors, to calculate the number of Twitter users identified in each county from 2014-2015. We calculate an estimated 2,390,091 Twitter users in Clinton-supported counties versus 402,202 in Trump-supported counties (among counties with at least one reported anti-Asian incident in 2020). Thus, while imperfect, this evidence suggests that the disproportionate spike in incidents in Trump-supported counties was not driven by greater exposure to his tweets in such counties.[27]

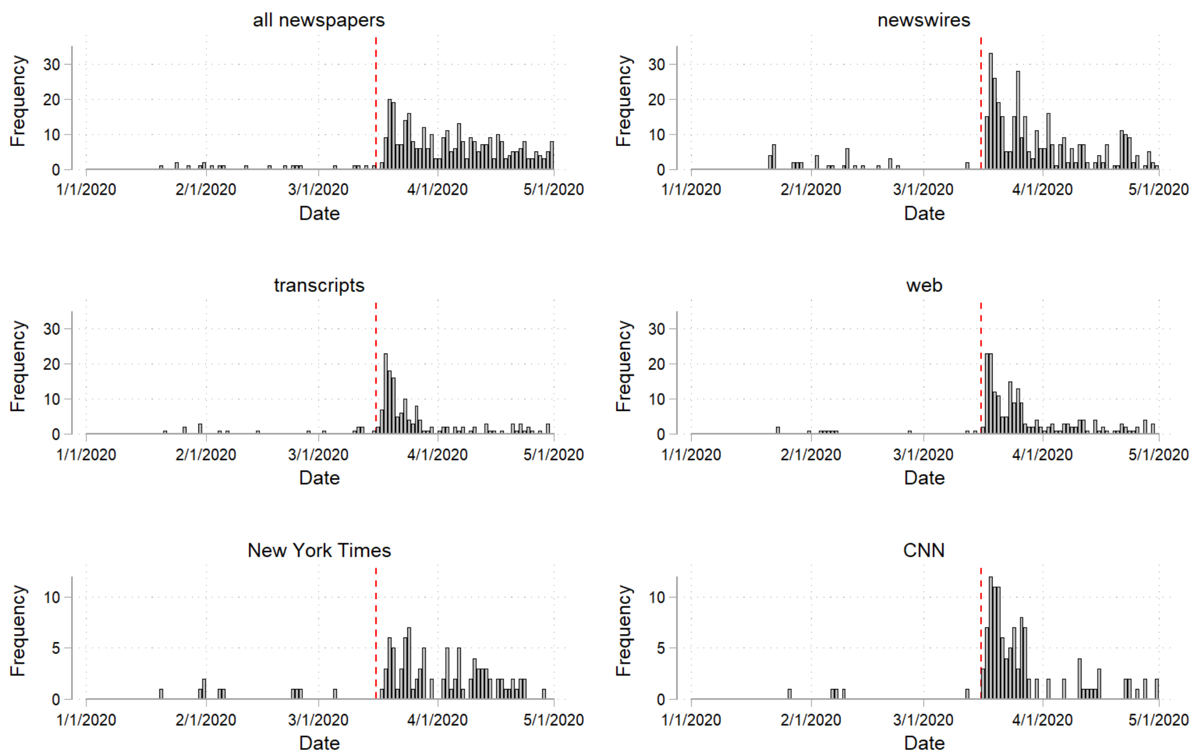
Along similar lines, we have also investigated how media outlets outside of Twitter may have contributed to propagating the effects we identify. We did so using Nexis Uni (formerly Lexis-Nexis Academic) to identify the number of media outlets using the phrase "Chinese Virus" each day. These data (depicted in Figure 2.5) show that it was very infrequently used in newspapers, newswires, TV/radio (wires), and online prior to Trump's tweets. They also show a massive increase immediately after Trump's tweets, which gradually declined in the following months. A similar trend is evident for use in the New York Times and CNN (Figure 2.5), which we interpret as evidence of widespread exposure to the phrase "Chinese virus" beyond Trump supporters.

One might also wonder if the apparent effects of Trump's "Chinese Virus" tweets resulted from a heightened media attention to those particular tweets relative to his other tweets. However, there are no apparent irregularities in the number media outlets mentioning both "Trump" and "Twitter"

[26] There is perhaps some evidence of an effect on the following day for incidents in the "other" category (Figure B.15, Panel C).

[27] We recognize that it is likely that per-user exposure to Trump's tweets is likely to be higher in Trump-supported counties than Clinton-supported counties. That said, Twitter users in Trump-supported counties would have to be 84 percentage points more likely to be exposed to his tweets than Twitter users in Clinton-supported counties in order to be more numerous than those exposed to his tweets in Clinton-supported counties.

Figure 2.5: Media Outlets Using the Phrase "Chinese virus"

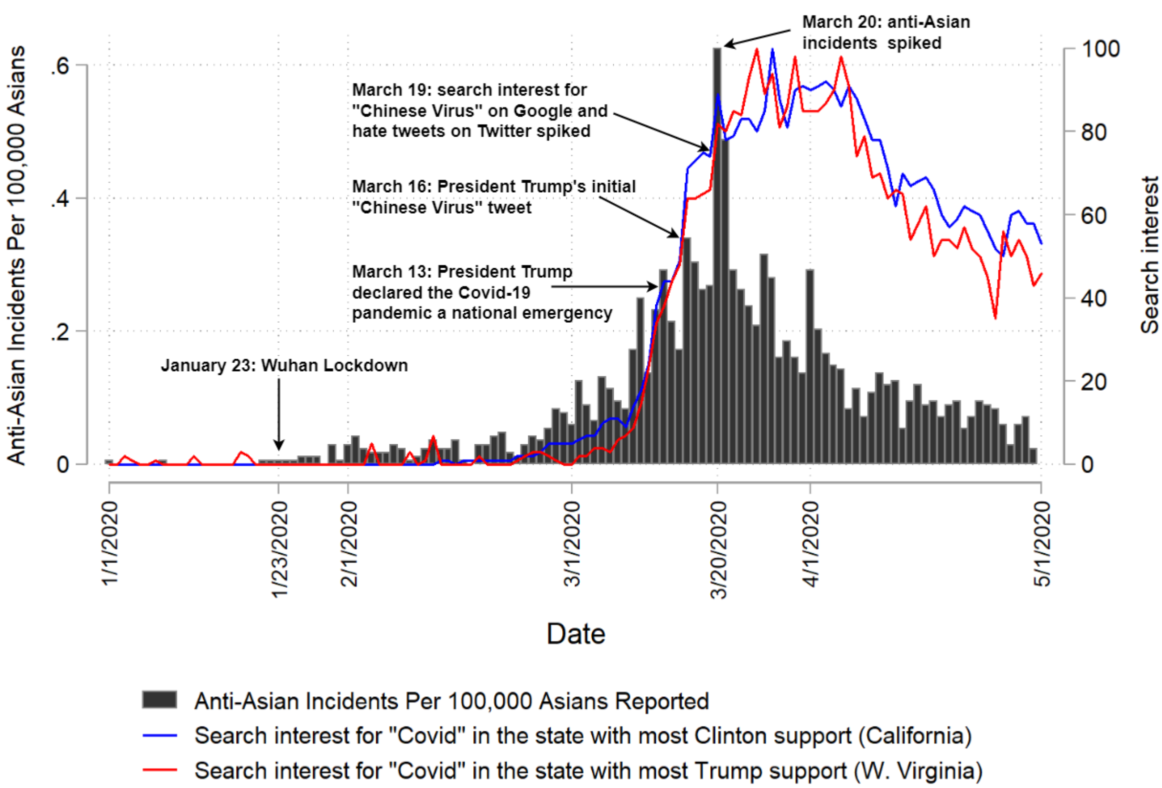


Notes: The red vertical lines are drawn the date of Trump's first "Chinese virus" tweet (3/16/2020). New York Times mentions include the newspapers only while CNN mentions include newswires, transcripts, and web.

around that time (Figure B.16). Along similar lines, the frequency with which Trump was tweeting was not irregular at that time (Figure B.17).

Closely related, but regarding a potential confounder rather than a mechanism, the differences we observed across Trump- and Clinton-supported counties could result from differences in concerns about COVID-19 across these counties. That said, as we noted above, these estimates are robust to the inclusion of county-day controls for COVID-19. Whereas that approach captures changes in local concerns to a degree, it leaves open the possibility that differences in more-general concerns about COVID-19—perhaps propagated by national media sources—may have changed differently for Trump- and Clinton-supported counties in a way that could explain the differences in described in the previous section. We investigated this possibility by evaluating search interest for "Covid" in the state where Trump received the greatest support in the 2016 presidential (West Virginia) versus search interest for "Covid" in the state where Clinton received the greatest support (California). As shown in Figure 2.6, search interest for "Covid" began to increase rapidly in both states before Trump's initial "Chinese Virus" tweet and increased somewhat more rapidly in California than West Virginia between the time of that tweet and the spike in anti-Asian incidents. Given these empirical regularities, we think it is unlikely that differences in general concerns about COVID-19 explain the spike in anti-Asian incidents in Trump-supported counties relative to that observed in Clinton-supported counties.

Figure 2.6: Trends in General Interest in "Covid"



Notes: Google Trends search interest data are standardized so that 100 represents peak search interest. This has been done separately for California and W. Virginia.

2.6 Limitations

While our findings strongly suggest that Trump's tweets resulted in increased anti-Asian behavior, the analyses are not without limitations. The nature of our empirical approach highlights a short-term spike in reported anti-Asian behavior. We are limited by statistical power in our capacity to speak to the longer-term effects of the president's rhetoric on anti-Asian sentiment or behavior. Specifically, while there is no detectible increase in anti-Asian behavior in Trump-supporting counties over the subsequent weeks, we cannot rule out the possibility of meaningful long-run effects.[28] Furthermore, it is possible that Trump's tweets had indirect and longer-term effects on anti-Asian behaviors outside of his core group of supporters. If stronger in Clinton-supporting counties, these indirect effects would lead us to underestimate the impact of his tweets over the longer term.

The data used in the paper, while novel, also have limitations. As with much self-reported crime data, we expect that there is significant underreporting of anti-Asian incidents to the STOP AAPI HATE reporting center. Assuming that this underreporting is uncorrelated with the effects we consider, estimated effects on the number of reports will be a conservative lower bound for the true effect on the number of incidents (whereas estimated percent changes will be accurate despite the measurement error).[29] However, there may be a concern that Trump's tweets influenced reporting of anti-Asian incidents independently of any effects on whether such incidents occurred. Similarly, individuals may have been more likely to report incidents after Trump's first "Chinese virus" tweet because the STOP AAPI HATE reporting center only began collecting data on March 19, 2020. That said, we note that our analyses focus on the date of the incident (and not of the report), and the data include a very large number of reports of incidents that occurred before this date (Figure 2.1). In addition, the spike in incident occurrences does not correspond to the date when the largest

[28]For example, the upper bound of the 95 percent confidence interval of the estimated effect over the period covering the 3rd to 28th week after Trump's initial tweet includes effects as large two-thirds of the mean level of anti-Asian incidents in Clinton supporting counties in the week before Trump's initial tweet. This estimate is produced using a straightforward difference-in-differences approach in which we drop the 1st and 2nd week after Trump's initial tweet.

[29]Authors' estimates from the National Crime Victimization Survey (NCVS) and FBI's Uniform Crime Report (UCR) hate crime statistics suggest that the number of hate crime incidents may be more than 30 times the reported number. Reports of incidents to the STOP AAPI HATE reporting center may understate the number of incidents by an even greater magnitude since the center is relatively new and many victims may be unaware of its existence.

number of reports were filed. Specifically, the data show a spike in incidents occurring on March 20 whereas more reports were filed on March 23 than were filed on March 20 or March 21, or on any other day in March or April of 2020 (Figure A18). Finally, we note that the relative spike in reported incidents in Trump-supporting counties is the opposite of what we would expect if it were driven primarily by an increase in reports resulting from increases in awareness of or sensitivity to anti-Asian incidents amongst more-liberal leaning individuals.

Another important limitation of our analysis is that we do not know the motivations of the actors or the exact mechanism by which Trump's tweets generate anti-Asian behavior. For example, the actors may be engaging in these acts because they believe their behavior is in service of a societal moral good that Trump has signaled (Ginges & Atran, 2009; Fiske & Rai, 2014) or the acts may be instances of a loss of self-control that the actors will regret (Baumeister & Heatherton, 1996; Card & Dahl, 2011; Dollard et al., 2013). Alternatively, the effects may represent an "emboldening effect" whereby individuals' determination of morally acceptable speech and behavior is influenced by the behavior they observe from elites, including tacit signals about what they condone (Newman et al., 2021). The effects might also be a result of "othering" whereby Trump's words heightened perceived differences in a manner that marginalized Asians (Gover et al., 2020; Reny & Barreto, 2022). Naturally, any of these effects may be amplified through the effects of peers (Sacerdote, 2014).

2.7 Conclusions

While there has been extensive media attention related to President Trump's rhetoric and influence, it is often difficult to separate his direct effect from underlying trends in behavior that coincide with his comments. We take advantage of new high-frequency data to demonstrate that his inflammatory remarks about COVID-19 resulted in a significant spike in anti-Asian behavior, with these effects concentrated in counties with greater support for the president, which is notable because these counties are disproportionately rural while the vast majority of Asian Americans live in urban areas. Google search data underscores the direct link between Trump's remarks, the rise in interest in the "Chinese Virus", and the spike in subsequent anti-Asian behavior.

Our findings provide empirical support for President Trump's capacity to influence not only the beliefs of his supporters, but also their actions. While a large body of work suggests that high-profile individuals can increase pro-social beliefs and behaviors, we demonstrate that they can have significant detrimental effects as well, even when the technology of social media substantially limits what they can say. This finding may have important implications given the recent rise of populist leaders pushing antisocial beliefs and behaviors on topics ranging from vaccine hesitancy to the treatment of immigrants.

3. WHO SELF-SELECTS INTO COMMITTEES: THE PRO-SOCIAL OR THE CORRUPT?

3.1 Introduction

[1]The management and redistribution of public resources often relies on a committee. Such governing body may consists of unpaid voluntary individuals or full time civil servant. For example, in Home Owners' Associations (HOAs), board members jointly make decisions on renovation and maintenance at the neighborhood level; in Parent-Teacher Organizations (PTO), elected parents work together with teachers to supplement the education experience of their children; finally, government officials at both the state and local level work together to provide people with public service. On the one hand, selection into such committees is costly. Data from the US Bureau of Labor Statistics show that federal workers on average earn 20% less than private sector workers with similar responsibilities. Moreover, many countries in the world requires extensive examination for entry into civil service (Sundell, 2014).[2] Finally, even for entry into part-time voluntary committees such as HOA, the election stage could costs a tremendous amount of time and effort. Thus, given the huge costs of entrance, one might expect that the prosocial and intrinsically motivated individuals are more likely to self-select into Committees (Besley & Ghatak, 2005; Serra et al., 2011; Besley & Ghatak, 2018; Ashraf et al., 2020; Bertrand et al., 2020). However, on the other hand, selection into committees can be extremely lucrative from a monetary perspective. It offers opportunities to embezzle public funds, especially in an environment where corruption is widespread and there is little transparency and accountability. Corruption from the government has been documented extensively in the literature (see Olken & Pande, 2012 for a review). In recent years, there has also been an alarming increase in cases of fraud and embezzlement in community based organizations[3] due to the absence of regulation.[4] Thus, given the huge monetary returns in

[1]This chapter is a joint work with Dmitry Rvykin (Florida State University) and Danila Serra. We thank Catherine Eckel for helpful comments.

[2]For example, in 2021 China, more than 2 million applicants registered for the nation wide civil service exam. Applicants have only a 1-in-68 chance of success.

[3]For example, leaders in Florida's biggest HOA are charged for $2 million fraud.

[4]Since association boards are largely unregulated by any state or federal agency, people can only turn to social media to seek solutions. For example, Reddit has a dedicated discussion board where more than 200,000 users share

unethical conduct, one might expect that the dishonest individuals who are more prone to corruption are more likely to self-select into committees. Despite the recent growth in the literature on selection into the public sector, results are mixed.[5]

In this paper, we employ a laboratory experiment to address the following research questions. First, who is more likely to self-select into committees? Is it the prosocial and intrinsically motivated individuals, or is it individuals who are more prone to corruption? Second, does self-selection depend on the status quo level of corruption? For example, are corrupt committees more likely to attract corrupt individuals? Given the huge costs and negative externalities of corruption, understanding who selects into service positions is extremely important. The answer to our research questions can not only help explain the variation in the levels of corruption across different organizations/countries, but also help improve the screening and recruitment procedure.

In light of our research questions, we need to develop a reliable measure of a committee's corruption level. In addition, we also need to measure the corruptibility of individuals who wish to join the committee. Thus, it is extremely challenging to answer our research questions using observational data. To circumvent data limitations and identification challenges, we design and run a two-stage laboratory experiment. The first part of the experiment consists of four one-shot social preference games, which we call, the *Pre-games*. We built upon the existing literature and carefully constructed these games to measure subjects' corruptibility along on two dimensions: 1) the degree of prosociality; 2) the degree of dishonesty. Then in the second part of the experiment, subjects participate in the *Committee Game*. The design simulates a society where a committee is in charge of managing a public fund and can choose to steal from the fund without the public knowing. Subjects are placed into groups of 8, with 5 Citizens and 3 Committee Members. For the Citizens, they participate in a simple real effort task each round to earn wages. However, Citizens must deposit a portion of their earnings into a public fund. The 3 Committee Members are in charge of managing the public fund. Depending on the Committee's performance and chance, the

stories on HOA-related problems.

[5]For example, Hanna and Wang (2017)'s study in India showed that college students with a higher propensity for cheating are more likely to enter the government upon graduation. Barfort et al. (2019)'s study in Denmark found the opposite.

public fund can either be tripled or lost. If the fund is tripled, Committee Members will jointly decide whether to redistribute the money equally among all members of the society, including the citizens, or embezzle the fund and only redistribute among the 3 of them. The game is played in 4 blocks of 10 rounds. At the beginning of a new block, 1 Committee Member will be randomly chosen to step down and replaced by a Citizen who are interested in joining the committee.

There are two features of our design that are worth noting. First, by construction, there is a lack of transparency from the Citizens' perspective. In the case that the Citizens receive 0 dividends from the public fund, it could be due to one of the three reasons: 1) the Committee Members failed their tasks; 2) The Committee Members completed their tasks but were unlucky; 3) The Committee Members completed their tasks but chose to embezzle the fund. Thus, the Citizens will not have perfect information and will have to form subjective beliefs about the Committee's action. Second, we implement different treatment conditions where we exogenously vary the status quo level of corruption. In the Honest (Corrupt) Committee Treatment, we place the subjects with the lowest (highest) corruptibility index into the initial Committee. Initial Committee Members are informed of the selection rule. The combination of these two features allows us to study whether and to what extent individuals' decision to join a committee depends on their own type (honest vs dishonest, and prosocial vs self-interested), and their subjective beliefs of how (dis)honest the existing Committee Members are.

During Fall 2022 and Spring 2023, we recruited 224 subjects from Texas A&M University's undergraduate student body to participate in our experiment. We have five main findings. First, using the Corruptibility Index constructed from the Pre-games, we are successful in creating corrupt vs honest initial Committees. The average Corruptibility Index of Committee Members is 3.33 in the Corrupt Committee Treatment, and 0.74 in the Honest Committee Treatment (p-value=0.00). Moreover, In the Corrupt Committee Treatment, the initial Committee Members are significantly more likely to vote for embezzlement, and the public fund is significantly more likely to be embezzled. However, this gap gradually shrinks over time and is eventually reversed in the last 10 rounds. Second, we show that despite the lack of transparency, Citizens are able to (almost) correctly form beliefs about the Committee's embezzlement behavior. Thus, we successfully generated environments with different levels of embezzlement where Citizens' beliefs about embezzlement closely align with the unobserved status quo level of embezzlement. Third, we find that Citizens in the Corrupt Committee Treatment express significantly higher interest to join the Committee compared to Citizens in the Honest Committee Treatment. This pattern is true regardless of the Citizen's corruptibility. This suggests that Citizens are more attracted to corrupt committees with higher status quo level of embezzlement. Fourth, we find evidence that Committee Members vote in line with their type in both treatment conditions, i.e., members with higher (lower) Corruptibility Index are more (less) likely to vote for embezzlement. This suggests that the more dishonest and corruptible players enter the Committee to steal from the public fund. In comparison, the more prosocial and honest type of subjects enter the Committee with the aim to reinforce the positive behaviors. Finally, we find evidence that the honesty trait and the prosocial trait play different role in determining Committee Members' decision. Results from chat analysis provide further support for this pattern.

Our paper contributes to two strands of literature. First, our paper contributes to the vast literature on selection into public service. The idea that people may sort into jobs based on their individual preferences and job attributes has been around for a long time. Over the years, scholars have investigated the role of mission-matching between individuals and organizations (Besley & Ghatak, 2005; Serra et al., 2011), self-image concerns (Bénabou & Tirole, 2011), financial incentives (Dal Bó et al., 2013; Deserranno, 2019), and career benefits (Ashraf et al., 2020) in shaping a person's occupational preference. In recent years, a growing body of papers have investigated the relationship between individual characteristics such as propensity for cheating and self-selection into public service. The results are mixed. Hanna and Wang (2017) recruited both university students and government workers in India to participate in their laboratory experiments and survey studies. They found that students who cheat more in a dice task and students who donate less to charities in a dictator game are more likely to prefer public sector jobs over private sector jobs. Moreover, they showed that cheating in the dice task predicts fraudulent absenteeism among government workers. A number of replication studies ensued, yet produced contrasting patterns. Barfort et al. (2019) used similar experimental design as Hanna and Wang (2017), but conducted their study in Denmark, the world's least corrupt country. They found a positive selection pattern where honest students who cheat less in a dice game exhibit stronger preference for government jobs. In a similar fashion, Friebel et al. (2019) showed that police applicants in Germany display higher trustworthiness than non-applicants, as measured by behaviors in the Trust Game. Finally, Gans-Morse et al. (2021) replicated Barfort et al. (2019)'s results with Russian students, despite the high level of corruption in Russia. The pattern is complicated even further by experimental studies with different designs (Banerjee et al., 2015; Banuri & Keefer, 2016; Brassiolo et al., 2021).[6] Given the discrepancy in results from studies conducted in low corruption countries and high corruption countries, it seems to suggest that in addition to individual characteristics, the existing level of corruption prevailing in the public sector matters as well. However, no study has examined the role of the individual's propensity for corruption and his/her subjective beliefs about corruption opportunity in one framework. Moreover, studies on career choices often fail to track subjects' behaviors once they start the job. For example, it is unclear whether prosocial individuals continue to behave in a prosocial manner or do they change over time. Our paper introduces a new variant of the embezzlement game. Our experiment mimics a common real-life scenario where Citizens are required to contribute to a collective fund managed by an administrative agency but do not have perfect information on how the agency operates. We extend the existing literature in two ways. First, Citizens in our experiment need to form their perception of how corrupt the Committee is based on available feedback each round. This allows us to observe how Citizens' beliefs affect their choice of joining the Committee. Second, the dynamic setting of our game allows us to observe what happens after a Citizen enters the committee. Our paper has important policy implications.

[6] Banerjee et al. (2015) conducted an embezzlement experiment with students from two Indian universities in the lab and showed that aspirant bureaucrats engage in more corrupt behavior than private sector aspirants. Banuri and Keefer (2016) used a series of real effort tasks that vary in payoff structure and found that university students in Indonesia who chose to work in public organization are significantly more prosocially motivated than a comparable sample of general workers. Brassiolo et al. (2021) who conducted a lab experiment where subjects are offered contracts that vary in opportunities of rent extraction. They found that the corrupt contract attracts dishonest individuals and repels honest ones.

Results from our experiment show that subjects show more interest in joining a corrupt committee regardless of their propensity for corruption. We also find evidence that subjects' voting behavior in the committee is in line with their "type". Taken altogether, our results highlight the importance of the screening process for public servants. A screening method that focuses on characteristics such as honesty and prosociality can be an effective way to reduce corruption.

Second, our paper broadly relates to the literature on corruption. While the exisiting literature on bribery is vast (Rose-Ackerman, 1975; Shleifer & Vishny, 1993; Abbink et al., 2002; Bertrand et al., 2007; Olken, 2007; L. Cameron et al., 2009; Barr & Serra, 2010; Banuri & Eckel, 2015; Salmon & Serra, 2017; Gneezy et al., 2019), embezzlement behaviors such as illegal appropriation of funds are largely overlooked in comparison. Some exceptions include experimental studies such as Azfar and Nelson Jr (2007) and Barr et al. (2009). They showed that increasing government wages reduces embezzlement and increases the quality of service. In addition, Attanasi et al. (2019) focuses on the psychological cost of embezzling and showed that subjects' decision to embezzle is influenced by others' expectations and guilt aversion. There are several field studies that examined whether increasing the probability of audit reduces embezzlement. Reinikka and Svensson (2005) showed that providing schools and parents with information on how to monitor local officials successfully reduced embezzlement and increased student enrollment. Olken (2007) conducted a field experiment in Indonesia villages and showed that increasing the probability of government audits significantly reduced missing expenditures. Di Falco et al. (2016)'s study in Tanzania showed that depending on the length of the transfer chain and the position of the intermediary, increasing information transparency could reduce embezzlement. Our experiment simulates a common situation where a committee has discretionary power over the redistribution of a public fund. Thus, in our setting, the decision maker is a group of Committee Members rather than an individual. While the existing experimental literature suggests groups tend to make more self-interested decisions (Charness & Sutter, 2012), no other paper has studied embezzlement in a group-decision environment to the best of our knowledge.[7] Our setting allows us to examine how

[7] On a related note, the experimental literature on peer effects highlights the contagion effect of anti-social behaviors (e.g., Gino et al., 2009; Dimant, 2019)

one's perception of the embezzlement activity of existing group members affect the selection into the group, as well as how one's behaviors evolve after having joined the group and working with group members with different propensity for embezzlement.

The remainder of this paper is structured as follows. Section 3.2 introduces our theoretical model and predictions. Section 3.3 describes our experimental design. Section 3.4 presents the results from our experiment. Section 3.5 concludes.

3.2 Model

In this section, we build a theoretical model to analyze how a Citizen's decision to enter the Committee depends on the Citizen's preference for dis(honesty) and prosociality, and the Citizen's belief about the status quo corruption level of the existing Committee.

3.2.1 The Setup

There is a society consisting of n Citizens, indexed by $i \in \{1, \ldots, n\}$, and k Committee Members, indexed by $j \in \{1, \ldots, k\}$. Each Committee Member receives a flat wage $w_c > 0$, and each Citizen i receives a wage $w_i = w_c + s_i$, where $s_i \geq 0$ are Citizens' earnings from an incentivized task.

Citizens pay a tax on their earnings at rate $\tau \in [0, 1]$. Total tax revenue, $T = \tau \sum_{i=1}^{n} s_i$, is transferred to the Committee which then invests it into a project with stochastic returns. With probability $p \in [0, 1]$, the project is successful and yields revenue rT, $r > 1$; and with probability $1 - p$ the project is unsuccessful, and the investment is lost.[8]

The outcome of the project is only observed by the Committee Members. If the project is not successful, Citizens receive their after-tax income, $w_c + (1-\tau)s_i$, and Committee Members receive w_c. If the project is successful, the Committee has two options: (i) to distribute the revenue equally among members of the society so that each Citizen i gets payoff $w_c + (1 - \tau)s_i + \frac{rT}{n+k}$, and the Committee Members get $w_c + \frac{rT}{n+k}$; and (ii) to embezzle the money, in which case Citizens get

[8]In the experiment, we have Committee Members perform a task. If their joint performance on the task reaches a certain threshold, the project is successful with probability p. If their performance falls short of the threshold, the project fails. For the purposes of this section, we will assume that Committee Members always reach the threshold. If there is some probability $\kappa > 0$ that they do not, the results in this section still hold with p replaced by $p(1 - \kappa)$.

their after-tax income $w_c + (1-\tau)s_i$, and each Committee Member receives $w_c + \frac{rT}{k}$.

We assume that, in the absence of embezzlement, investment is efficient in expectation; that is, all Citizens' expected payoffs exceed their pre-tax wages:

$$(1-\tau)s_i + \frac{prT}{n+k} > s_i, \quad i = 1, \ldots, n.$$

Thus, we require that

$$\frac{prn\bar{s}}{n+k} > s_{\max}, \tag{3.1}$$

where $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$ is average earnings, and $s_{\max} = \max\{s_1, \ldots, s_n\}$ is the largest earnings. For example, if all Citizens earn the same amount, condition (3.1) is simply $pr > \frac{n+k}{n}$.

Each agent is characterized by a pair of parameters (m, b), where m is a moral cost measuring how much the individual dislikes being involved in embezzlement; and b is a prosociality parameter measuring how much the individual values the money transfer to Citizens. We will sometimes refer to vector (m, b) as the agent's *type*. Let $Z \in \{0, 1\}$ denote an indicator equal to 1 if money from a successful investment is embezzled and zero otherwise. If the project is successful, the utility of Committee Member j is given by

$$\begin{aligned} u_j &= w_c + Z\left(\frac{rT}{k} - m_j\right) + (1-Z)\left(b_j + \frac{rT}{n+k}\right) \\ &= w_c + Z\left(\frac{nrT}{k(n+k)} - m_j - b_j\right) + b_j + \frac{rT}{n+k}. \end{aligned}$$

Thus, Committee Member j personally prefers embezzlement if $m_j + b_j < \frac{nrT}{k(n+k)}$.[9] As expected, the higher the moral cost of embezzlement and the higher the prosociality parameter, the less likely a Committee Member prefers embezzlement. The two parameters serve as substitutes and can be combined into one intrinsic motivation parameter $a = m + b$. For brevity, we will sometimes refer to a as the agent's type as well. Thus, members with $a_j < (>) a_e \equiv \frac{nrT}{k(n+k)}$ prefer embezzlement (honesty).

[9] Here and in what follows, to simplify matters, we ignore the possibility of indifference. It happens with probability zero if type parameters have an absolutely continuous distribution in the population.

3.2.2 Voting on the Committee

When the project is successful, the Committee makes a decision whether or not to embezzle the revenue by majority voting. There are multiple voting equilibria. For example, if $k = 3$ and two Committee Members vote in favor of embezzlement, any mixed voting strategy by the third member is a best response. To fix matters, we restrict attention to equilibria with *sincere voting*, such that members vote according to their preferences. For k odd, the outcome is determined by a member with the median type $a_{\text{med}} = \text{median}\{a_1, \ldots, a_k\}$. Embezzlement happens when $a_{\text{med}} < a_e$.

3.2.3 Citizens' Decision to Join the Committee

The stage game described above is repeated for some number of periods, after which one randomly selected member of the Committee steps down. Consider a Citizen with type (m, b), the corresponding combined type a, and earnings s, deciding whether she should join the Committee.

Let $\mu(a)$ denote the Citizen's belief that *after she joins the Committee*, the Committee will be corrupt. Further, let $\tilde{\mu}$ denote the belief that the Committee will be corrupt if instead someone else joins it. The Citizen prefers to join the Committee if

$$w_c + p\left[(1 - \mu(a))\left(b + \frac{rT}{n+k}\right) + \mu(a)\left(\frac{rT}{k} - m\right)\right] > w_c + (1 - \tau)s + p(1 - \tilde{\mu})\frac{rT}{n+k}.$$

This condition can be transformed as

$$p\big[b + \mu(a)(a_e - a)\big] > (1 - \tau)s - \frac{p\tilde{\mu}rT}{n+k}. \tag{3.2}$$

The right-hand side of (3.2) is the Citizen's net expected utility if someone else joins the Committee. The first term is the net gain in wages, and the second term is the expected loss from embezzlement.

The left-hand side of (3.2) is the Citizen's net expected utility from joining the Committee. A Citizen may want to join the Committee for one of two reasons: (i) to experience utility from public service, i.e., from being in charge of transferring money to Citizens; and (ii) to earn money from embezzlement. These two motives are represented by the two terms inside the square brackets. The first term, b, is the utility gain, for a prosocial Citizen, from the ability to provide the service to other Citizens. The second term, $\mu(a)(a_e - a)$, is the net gain from embezzlement, which is positive for a Citizen with $a < a_e$, and negative otherwise. Note that there is a trade-off between the two terms: As the Citizen's a goes up, the second term decreases and eventually becomes negative; concurrently, the first term will increase if the increase in a is due to prosociality. If, however, the increase in a is solely due to aversion to corruption, then the first term will not increase, and the Citizen will be less likely to join the Committee.

The latter effect is due to the moral cost of embezzlement the Citizen suffers from being a member of a corrupt Committee, even if she does not support corruption. In this case, staying out of the Committee allows the Citizen to avoid being involved in "dirty politics." An alternative explanation is that the Citizen stays out in order to avoid the feeling of disappointment or regret from not being able to provide services to the community.

3.2.4 Calculating $\tilde{\mu}$ and $\mu(a)$

For simplicity, assume that Citizens are naïve in their Committee entry decisions. That is, each Citizen makes her entry decision on the basis of condition (3.2), but she does not take into account that other Citizens do so as well. Under this assumption, from a Citizen's viewpoint, the assignment of others' types on the Committee is random.[10] Suppose all agents' types are i.i.d. with a joint distribution $F(m, b)$. Let $F_a(\cdot)$ denote the resulting distribution of intrinsic motivation parameter $a = m + b$.

Assuming k is odd, a random Committee is corrupt if $\frac{k+1}{2}$ or more of its members are. This gives

$$\tilde{\mu} = \sum_{j=\frac{k+1}{2}}^{k} \binom{k}{j} F_a(a_e)^j [1 - F_a(a_e)]^{k-j}.$$

[10] In the full entry equilibrium, sophisticated Citizens take into account that other Citizens also decide whether or not to enter on the basis of condition (3.2). This will affect the calculation of $\mu(a)$ and $\tilde{\mu}$, in which the distribution of other Committee Members' types will be endogenously updated.

To find $\mu(a)$, consider two cases. If $a > a_e$, at least $\frac{k+1}{2}$ others have to be corrupt; whereas, if $a < a_e$, it is sufficient to have $\frac{k-1}{2}$ other corrupt members. This gives

$$\mu(a) = \sum_{j=\frac{k+1}{2}}^{k-1} \binom{k-1}{j} F_a(a_e)^j [1 - F(a_e)]^{k-1-j} + \binom{k-1}{\frac{k-1}{2}} F_a(a_e)^{\frac{k-1}{2}} [1 - F_a(a_e)]^{\frac{k-1}{2}} 1_{a<a_e}.$$

3.2.5 Beliefs about the Committee Being Corrupt

In this section, we discuss how Citizens update beliefs about whether or not the Committee is corrupt by observing the history of investment outcomes. To simplify matters, we assume that (i) the Committee's strategy is stationary; and (ii) Citizens are Bayesian updaters.

Let μ_c denote the Citizens' prior belief that the Committee is corrupt. Suppose Citizens have played the game for t periods and in t' out of those t periods they received zero returns on their tax investment. Let $\mu_{t',t}$ denote the resulting posterior belief that the Committee is corrupt.

For illustration, we will first consider a simple setting where Committee Members are non-strategic, in the sense that they do not try to manipulate Citizens' beliefs. In this case, a corrupt Committee always embezzles, whereas an honest Committee is always honest. The first observation is that if $t' < t$, then $\mu_{t',t} = 0$. In other words, it is sufficient for Citizens to observe a positive return once to conclude that the Committee is honest. If, however, all t periods produced zero returns, the posterior probability that the Committee is corrupt is

$$\mu_{t,t} = P(C|t) = \frac{P(t|C)P(C)}{P(t|C)P(C) + P(t|NC)P(NC)}.$$

Here, $P(t|C)$ is the probability of observing t losses when the Committee is corrupt, which under our assumption of nonstrategic Committee is equal to one; and $P(t|NC)$ is the probability of observing t losses if the Committee is honest, which is equal $(1-p)^t$. This gives

$$\mu_{t,t} = \frac{\mu_c}{\mu_c + (1-p)^t(1-\mu_c)}.$$

As expected, as t increases the posterior is approaching one. For example, for $p = 0.7$ and $\mu_c = 0.5$ we have

$$\mu_{1,1} = 0.769, \quad \mu_{2,2} = 0.917, \quad \mu_{3,3} = 0.974, \quad \mu_{4,4} = 0.992.$$

Thus, already by period 3 Citizens are almost certain the Committee is corrupt. The speed of this updating process is much more sensitive to the probability of project success, p, than to the prior, μ_c. For example, if $p = 0.3$, with the same prior Citizens would have $\mu_{5,5} = 0.856$ and $\mu_{9,9} = 0.961$. However, with $p = 0.7$ but prior $\mu_c = 0.2$ they would still have $\mu_{5,5} = 0.990$.

For this reason, it may be in the Committee Members' interest (if they are corrupt) to manipulate Citizens' beliefs by sometimes pretending to be honest. We will now consider the behavior of such a strategic Committee. Suppose a corrupt Committee decides to pretend they are honest with probability α. That is, if a project is successful, with probability α the Committee will not embezzle. In this case, the probability of a loss happening in any given period if the Committee is corrupt is $1 - p\alpha$. This gives the posterior belief after having observed t' losses:

$$\begin{aligned}\mu_{t',t} = P(C|t') &= \frac{\binom{t}{t'}(1-p\alpha)^{t'}(p\alpha)^{t-t'}\mu_c}{\binom{t}{t'}(1-p\alpha)^{t'}(p\alpha)^{t-t'}\mu_c + \binom{t}{t'}(1-p)^{t'}p^{t-t'}(1-\mu_c)} \\ &= \frac{(1-p\alpha)^{t'}\alpha^{t-t'}\mu_c}{(1-p\alpha)^{t'}\alpha^{t-t'}\mu_c + (1-p)^{t'}(1-\mu_c)}.\end{aligned}$$

Table 3.1 shows the dependence of $\mu_{t',t}$ on t' using the parameters from the previous example for $t = 9$ and three different values of α.

As expected, the posterior belief increases in t' for each α. Interestingly, it can be nonmonotone in α for some values of t'. For low t', the belief increases in α, meaning that when few losses are observed, it is more likely that the Committee is corrupt the more likely it pretends to be honest. When t' is high, the dependence is reversed: now it is less likely that the Committee is corrupt the more likely it pretends to be honest. For $t' = 6$, which is large but not too large, it is more likely the Committee is corrupt when it pretends half of the time than when it pretends 30% of the time or 70% of the time.

We can think of the Committee choosing α to manipulate Citizens' beliefs, in a manner similar

Table 3.1: Posterior Beliefs after Observing t' Losses in t Periods

t'	$\mu_{t',t}$		
	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$
0	0.000	0.002	0.039
1	0.000	0.008	0.089
2	0.002	0.035	0.192
3	0.013	0.137	0.366
4	0.105	0.408	0.584
5	0.506	0.749	0.773
6	0.900	0.928	0.892
7	0.988	0.982	0.953
8	0.999	0.996	0.980
9	1.000	0.999	0.992

Notes: Parameters: $p = 0.7$, $\mu_c = 0.5$, $t = 9$

to information design and Bayesian persuasion. If the Committee's utility depends on Citizens' beliefs—for example, if the Committee values money but also values being perceived as honest—there is a trade-off, and we can characterize the "optimal" α.

3.3 Experimental Design

Given the nature of our research questions, it is extremely hard to answer them using observational data for two reasons. First, it is challenging to obtain an available and accurate measure of embezzlement in the field. As with corruption data in general, embezzlement is largely not observable due to its illegal nature.[11] Thus, our empirical strategy is restricted by the scarcity of available data. Second, to study how self-selection depend on the status quo level of corruption, we need to generate exogenous variation in the level of Committee's embezzlement activity. This is unfeasible from both a design and an ethical perspective. To circumvent the challenges from data and identification, we use a laboratory experiment. Our experiment proceeds in two stages, a series of short one shot pre-games followed by the main game, which lasts for 40 rounds. We

[11]Corruption data is usually collected from either government level audit or perception surveys (Ortiz-Ospina & Roser, 2016). None fits the purpose of our research

introduce the pre-games in Section 3.3.1, and the main game in Section 3.3.2.

3.3.1 The Pre-games

We call the first part of the experiment the Pre-games, which is made up of four one-shot social preference games. The purpose of the Pre-games is to measure subjects' preferences along two dimensions, 1) self-interest vs prosociality ; 2) honesty vs dishonesty. These two dimensions correspond to the b parameter and the m parameter introduced in our model section respectively. In the context of a committee with discretionary power over the distribution/embezzlement of a public fund, the b parameter characterizes a Committee Member's utility gain from providing public service to Citizens. The m parameter characterizes a Committee Member's utility loss (moral cost) from being involved in embezzlement. We then construct an overall index of corruptibility and characterize individuals into "types". This index serves as a proxy for the a parameter. In the second part of the experiment, we use the index to assign subjects to groups and roles. Our goal is to divide all experimental sessions into societies with an honest committee, and societies with a corrupt committee. In this subsection, we describe each Pre-game in detail.

3.3.1.1 Pre-Games: Giving VCM and Donation Game

We conceptually decompose an individual's propensity for prosocial behaviors into two components. The first component is the individual's willingness to coordinate and cooperate with others to solve collective action problems and achieve the socially optimal outcome. In order to capture this component, we employ a binary giving VCM game adapted from Barr et al. (2014). In this game, subjects are randomly matched with three other subjects in the session. Subjects are not informed of the identity of their matches. Each group member receives an endowment of 10 ECU. Subjects are told that each member of the group will have to independently decide whether to invest the 10 ECU in a Private account or a Group account. The total amount invested in the Group account will be multiplied by 1.6 and redistributed equally among each group member. Thus, if a subject invests the money in the Private account, he will earn $10 + 4 * N$ where N is the total number of people who invested in the Group account. Otherwise, the subject will earn $4 * N$. To help subjects understand the game and reduce their cognitive burden, we provide a table that shows the payoff for every possible scenarios. In addition, to make sure that subjects clearly understand the trade-off between maximizing individual payoff and maximizing group payoff, the table shows both the amount the subject will earn and the amount the group as a whole will earn. We chose this variant of the VCM game for three reasons. First, individuals' behaviors in the public goods game have been shown to be a good predictor of their behavior in everyday life. For example, Albanian parents who behave cooperatively in the public goods game are more likely to participate in school in school accountability institutions and national elections (Barr et al. (2014)). In addition, fishermen who behave cooperatively in the public goods game are less likely to exploit common pool resources (Fehr & Leibbrandt, 2011), and have higher productivity when fishing in teams (Carpenter & Seki, 2011). Second, the context of VCM closely resembles the committee setting our paper focuses on. Conceptually, a Committee Member has discretionary power over the distribution of public resources and will have to make decision on how to allocate the resources. Thus, the decision problem a subject is presented with in our VCM game simulates the committee setting. Third, we chose a binary version of the game over a standard version because it is quick and easy to implement. We do not want subjects to spend too much time on the Pre-Games because it might confound their behaviors in the main game.

The second component of an individual's propensity for prosocial behaviors is the individual's willingness to benefit others at the expense of self-interest. This component is relevant to our committee setting since the provision of public service often comes at high opportunity costs. To capture this component, we employ a donation game adapted from the classic Dictator Game (See Engel, 2011 for a review). In this game, subjects are given 20 ECU as initial endowment. Subjects are then provided with a list of seven charity organizations and are asked whether they want to keep the 20 ECU to themselves, or donate it to one of the seven charities.[12] The Dictator Game with charities as recipient is the most well-received experimental measure for individual prosociality.

[12]The charties presented are: The Texas A&M Foundation, The Salvation Army of Bryan/College Station, The Ronald McDonald House of Central Texas, Brazos Vallev Food Bank, Aggieland Humane Society, Bryan/College Station Habitat for Humanity, Doctors Without Borders

Moreover, existing studies has shown that there is a strong negative correlation between donation behavior and the individual's propensity for cheating, which is known to predict corruption (Hanna & Wang, 2017; Barfort et al., 2019). One concern is subjects might view the charity organizations presented to them as corrupt and incompetent. In that case, subjects will likely keep the money to themselves. Thus, in order to accurately capture subjects' propensity for prosocial behavior, we must make sure that subjects believe the charities are doing good work. We picked well-known charity organizations with good reputation such that our list consists of local organizations, state level organizations, and international organizations.

3.3.1.2 Pre-game: Taking VCM and Coin Game

We conceptually decompose an individual's propensity for dishonest behaviors into two components. The first component is the individual's willingness to maximize personal gains at the expense of the society. In the context of a committee managing public resources, this can be interpreted as appropriating resources for personal interest. In order to capture this component, we employ a binary VCM game with a taking frame adapted from Falk and Fischbacher (2002). This game is similarly structured as the VCM game with a giving frame introduced in the last subsection. Subjects are told that they are randomly matched with three other anonymous participants. In reality, subjects stay in the same group as the giving VCM game. Each group member will receive an endowment of 20 ECU. In addition, there is a Group account that contains 80 ECU. Subjects are asked to choose from two actions, take 10 ECU from the Group account, or do not take. If a group member chooses to take from the Group account, the account will be reduced by 20 ECU. After all group members have made their decisions, the amount that is not taken will be equally distributed among them. Thus, if a subject choose to take from Group account, he will earn $25-5*N$ where N is the total number of people who took from the Group account. Otherwise, the subject will earn $20-5*N$. To help subjects understand the game, we provide a table that shows the payoff for all possible scenarios. Similar to the giving VCM game, the payoff table shows both individual payoff and group-level payoff to highlight the trade-off. The biggest difference between this game and the giving VCM game is that this game has a taking frame, i.e., subjects are asked whether they want to take from the Group account instead of give (See Cartwright, 2016 for discussion on framing effects). We chose this variant for two reasons. First, existing literature has shown that public goods games with a taking frame are more likely to cause extreme behaviors while the giving frame is more likely to cause cooperative behaviors (Andreoni, 1995; Park, 2000; Cox & Stoddard, 2015; Khadjavi & Lange, 2015). Thus, the taking VCM game fits our purpose of simulating the misappropriation of public resources. Second, the wording and the setting in this game is similar to the giving VCM game except for the framing. Thus, this minimizes subjects' cognitive burden and reduces possible bias in the subsequent main game.

The second component of an individual's propensity for dishonest behaviors is the individual's willingness to lie for monetary gains. In the context of a committee managing public resources, this could be interpreted as sending false information to the public and lying about the availability of resources. To capture this component, we employ a variant of the coin toss game adapted from Abeler et al. (2014). Subjects are told that there is a coin and a tray on their desk.[13] We ask subjects to toss the coin 15 times in private and report the number of times they got tails. For each reported tail, subjects earn 1 ECU. Thus, the maximum earning from this game is 15 ECU. Since there is no way for us to verify the actual number of tails, subjects have an incentive to lie.[14] Many studies have shown that cheating behavior in variants of the coin toss game is correlated with real life unethical behaviors in a variety of settings. For example, illegal drug possession from prison inmates (Cohn et al., 2015), misconduct from high school students and university students (Cohn & Maréchal, 2018; Dai et al., 2018), not paying for public transportation (Dai et al., 2018). Most importantly, Hanna and Wang (2017) showed that cheating behavior in the coin toss game predicts corrupt behavior from health workers in the Indian government. Thus the coin toss game serves our purpose well. One concern is a high number of tails could simply due to luck and thus does not reflect cheating. We chose 15 because this number is sufficiently high to detect cheating. For example, a subject who reports 12 or more tails out of 15 tosses would be cheating with probability

[13] The purpose of the tray is to reduce the noise of coins hitting the desk and prevent coins from rolling around on the desk.

[14] Although an experimenter is present in the room, each subject's workstation is covered to protect privacy.

98%. While a higher number can achieve higher accuracy, it is more likely to raise suspicion from subjects.

3.3.1.3 Implementation

The instructions for each pre-game is provided on subjects' screen. Subjects are told that one game will be randomly selected for payment at the end of the experiment. Since the Main Game is conducted immediately after all subjects complete the Pre-Games, one might be concerned that subjects' response in the Main-Game is subjective to bias. For example, subjects' experience in the Pre-Games could have a spillover effect on the remainder of the experimental session. We take several measures to reduce bias and improve the quality of our data. First, subjects are not provided with any feedback on the Pre-Games until the end of the experiment. Thus, it is unlikely that subjects will be affected by their performance in the Pre-Games. Second, we used neutral names for each game. For example, instead of calling it the coin toss game or the cheating game, the game is displayed as "Activity Orange". This eliminates the possibility that subjects might be able to figure out the purpose of each game. Third, to control for ordering effects, the 4 games are presented to each subject in randomized order. Fourth, the identity of subjects is kept anonymous throughout out the entire experimental session. Therefore, although subjects might be matched with the same subjects as in the Pre-Games, it is unlikely to cause bias in subjects' behavior. Finally, although subjects are informed that the entire experiment consists of two parts, they are not aware that the two parts are connected. Therefore, it is unlikely that subjects will act strategically in the first part to control the second part of the experiment.

3.3.1.4 Creating the Corruptibility Index

The purpose of the Pre-Games is to create an overall measure of corruptibility, which we call the Corruptibility Index. We use subjects' response to all four pre-games to calculate the index as follows. First, in the Giving VCM Game, we assign 1 point for investing in the Private Account. Second, in the Donation Game, we assign 1 point for keeping the money. Third, in the Taking VCM Game, we assign 1 point for taking from the Group Account. Finally, in the Coin Game, we assign 1 point for reporting more than or equal to 12 tails. We assign 0 point otherwise. We then calculate the Corruptibility Index as the sum of the points scored. Thus, the maximum value of the index is 4, indicating the subject being the most corrupt. The minimum value of the index is 0, indicating the subject is the least corrupt. Subjects are not aware of this process.

3.3.2 The Committee Game

In the second part of the experiment, subjects participate in the *Committee Game*. Subjects are placed into groups of 8 which represents a society. 5 subjects are assigned the role of Citizen and the remaining 3 subjects are assigned the role of Committee Member. Our goal is to simulate an environment where members of a committee are put in charge of the redistribution of a public fund, which is generated by private Citizens. Committee Members can choose to secretly embezzle the funds without Citizens knowing. The game lasts for 4 sequences of 10 rounds. Subjects' role remain fixed for one entire sequence. At the beginning of a new sequence, 1 Committee Member is randomly chosen to step down, and replaced by a Citizen who expresses interest in joining the Committee. The combination of the Pre-Games and the Committee Game allows us to investigate two potential determinants of a Citizen's selection into the committee: 1) the Citizen's own corruptibility level; 2) the Citizen's belief about how corrupt the incumbent committee is. In this subsection, we describe the Committee Game in detail.

3.3.2.1 Citizens

The five Citizens engage in real-effort tasks and are required to contribute part of their earnings to the public fund. Depending on the action of the Committee Member and on luck, the public fund could be triple and Citizens could receive back an equal share of the fund. In addition, Citizens have the chance to periodically self-select into the committee.

Citizens' activity in the game consists of three parts. First, in each round, Citizens start with a fixed wage of 100 ECU. In addition, they are presented with three simple real-effort tasks and are given 30 seconds to complete the tasks.[15] If Citizens solve all three tasks correctly, they will

[15]Our choice of real-effort tasks is inspired by Charness et al. (2018). See Appendix C.1 for screenshots.

generate 50 ECU as additional earnings. However, Citizens can only keep 36% of the additional earnings to themselves, which is 18 ECU. The remaining 64% of Citizens' additional earnings, i.e., 32 ECU per Citizen, will be deposited into a public fund. Since there are 5 Citizens in a society, Citizens can generate up to 160 ECU to be deposited in the public fund. After all Citizens complete the tasks, they are given feedback on the number of tasks they solved correctly, as well as how many other Citizens solved three tasks correctly. Thus, Citizens are aware of the total amount of money in the public fund. Next, Citizens are told that the three Committee Members are in charge of the public fund. If Committee Members failed to complete their tasks, the public fund will be lost and neither the Citizens nor the Committee Members will receive any dividends. If Committee Members are successful, the public fund will be tripled with probability 80% or lost with probability 20%. If the public fund is successfully tripled, the 3 Committee Members have the task of redistributing the money equally among themselves and the Citizens. However, they can instead jointly decide to keep the money and divide only among the 3 of them. Thus, Citizens could either receive an amount equal to 1/8, or an amount equal to 1/3 of the total tripled money in the fund. The outcome will be decided by the 3 Committee Members via majority voting. While Citizens are informed of their earnings from the public fund, they do not observe the committee's task or the committee's action.

Second, in the end of round 5 and round 10 of every sequence, Citizens are required to report their opinions about the committee's performance. Specifically, Citizens see a summary table that shows the amount they received from the public fund for each of the past 5 rounds. We ask Citizens to report the number of times Committee Members decided to keep the money for themselves. Specifically, Citizens see the below question on their screen.

> Think about the X rounds when you did not receive money from the public fund. How many times do you think the public fund was successfully tripled and the Committee Members decided to keep the money?

To improve the accuracy of Citizens' beliefs, we incentivize the belief elicitation as follows. At the end of the experiment, one sequence will be randomly selected for payment. Citizens will receive 10 ECU for each correct guess in that sequence. Thus, the maximum earning Citizens can make from this task is 20 ECU. The belief elicitation serves two purposes and is crucial to answering our research questions. First, due to the lack of transparency, it is entirely possible that Citizens are unaware of the embezzlement behavior from the committee. Eliciting Citizen's beliefs allows us to examine whether Citizens' update their beliefs in response to the committee's action. It also allows us to examine the accuracy of the beliefs. Second, the belief elicitation allows us to answer one of our primary research questions: does self-selection depend on the status quo level of corruption.

The third part of the Citizens' activity is the selection into the committee. At the beginning of a new sequence, Citizens are told that one Committee Member will be randomly chosen to be replaced by a Citizen. Citizens are then asked whether they would like to join the committee. A new Committee Member will be randomly chosen from the Citizens who chose "yes". If no Citizen expressed interested, then the replacement will be randomly chosen among all eligible Citizens.[16]

There are several features of the design that are worth noting. First, the real-effort task for Citizens changes in every sequence. Thus, Citizens could participate in four different real-effort tasks in the entire game. The purpose of this change is to keep Citizens stimulated and engaged such that Citizens will not simply choose to join the committee because of boredom. Second, all real-effort tasks are designed to be easy such that there is always a sizeable amount in the public fund. In addition, Citizens are given the change to practice the task at the beginning of every sequence. Performance during the practice will not affect Citizens' income. Third, only the Committee Members will know if they were successful in the task, and whether the public fund got tripled or if the money is lost. In other words, if Citizens do not receive dividends from the public fund, it could be due to one of the three reasons: 1) Committee Members failed their tasks; 2) the public fund was lost due to poor luck; 3) Committee Members decided to keep the money in the fund for themselves. This lack of transparency forces Citizens to generate beliefs about the committee's action and thereby allows us to investigate how beliefs affect self-selection. Fourth,

[16]Citizens who served in the Committee before are not allowed to join again.

we parameterized the game such that Condition 3.1 is satisfied, i.e., investment is efficient in the absence of embezzlement. Since Citizens' expected payoffs exceed their pre-tax wages, rational Citizens will not choose to shirk. Finally, at the end of every round, we provide Citizens with a table that summarizes their earnings in that round.

3.3.2.2 Committee Members

The three Committee Members are in charge of managing the public fund. Depending on the Committee's performance and chance, the public fund can be either tripled or lost. If the fund is tripled, Committee Members will jointly decide whether to redistribute the money equally among all society members, or embezzle the fund and only redistribute among themselves.

The three Committee Members' earn a fixed wage of 80 ECU in each round. Their activity consists of four parts. First, they engage in a simple task that involves answering one general knowledge quiz question. If none of the Committee Members answered the question correctly, the public fund is lost. If at least one Committee Member answered the question correctly, the public fund will be either tripled with probability 80%, or lost with probability 20%. Second, in the case when the public fund gets tripled, Committee Members will have to vote for one of the two options: 1) Redistribute the money equally among every member of the society, i.e., the 3 Committee Members and the 5 Citizens will each receive 1/8 of the share; 2) Keep the money and redistribute it equally only among the 3 Committee Members, i.e., each Committee Members receive 1/3 of the share while Citizens receive nothing. The outcome is decided by majority voting. Third, the Committee Members are able to talk with each other via chat for 2 minutes at the beginning of round 1 and again at the beginning of round 6 of every sequence. We did not explicitly tell the Committee Members what they should discuss. Fourth, at the beginning of a new sequence, one Committee Member will be randomly chosen to be replaced by a Citizen. The chosen Committee Member will play the role of Citizen for the remainder of the game.

There are several features of the design that are worth noting. First, we intentionally set the Committee Member's fixed wage to be lower than Citizen's. This reflects the wage gap between the public sector and the private sector. Thus, there is a clear monetary cost in serving in the committee.

Second, the committee's task is designed to be easy such that at least one Committee Member should be able to complete the task correctly. Thus, the public fund will always be tripled with probability 80%. This ensures that Committee Members have to choose between redistribution and embezzle for the majority of the game. However, Citizens are not aware of what the tasks are. Third, the timing of the first chat coincides with Citizens' practice round. The timing of the second chat coincides with Citizens' belief elicitation. This reduces boredom from waiting for the other role. In addition, since all players will be typing, it reduces the likelihood that Citizens find out who the Committee Members are.[17] Fourth, we programmed the experiment such that it is always the oldest Committee Member who gets replaced by a Citizen. Thus, the entire committee will be replaced by the last sequence of the game. This generates more variation in the committee's composition. Fifth, if no Committee Member completes the task correctly, there will be no voting stage. Otherwise, the voting stage happens before they see how much money is in the public fund and whether it is tripled. This ensures that Committee Members' decision to embezzle is not affected by the size of the fund.

3.3.3 Treatment Conditions and Role Assignment

One of our key research questions is how the status quo of corruption affects the self-selection into committees. Therefore, it is crucial that we create societies with corrupt committees and societies with honest committees. In order to create variation in the committees' embezzlement activity, we implement two treatment conditions where we exogenously manipulate the composition of the initial committee.

As described in Section 3.3.1, we use subjects' response in the Pre-Games to create an overall measure of corruptibility. At the start of the Committee Game, we assign groups and roles based on the Corruptibility Index of all subjects who are present in the session. This proceeds in two steps for a 16-people session. First, after all 16 subjects have completed the Pre-Games, the program sorts all subjects by the value of their Corruptibility Index and assigns each subject an id based on the ranking. Thus, the three subjects with the lowest Corruptibility Index will be labeled as subject

[17] Since our session always have 2 societies, this likelihood is even smaller.

1, subject 2, subject 3. The three subjects with the highest Corruptibility Index will be labeled subject 14, subject 15, subject 16. Second, the program places subject 1, 2, and 3 into the initial committee of group 1, and places subject 14, 15, and 16 into the initial committee of group 2. We will call group 1 the Honest Committee Treatment, and group 2 the Corrupt Committee Treatment throughout the rest of this paper. In the final step, the remaining 10 subjects are randomly shuffled and assigned the role of Citizen in each group.

There are several features of the design that are worth noting. First, the only difference between the Corrupt Committee Treatment and the Honest Committee Treatment is the initial committee assignment. Second, our session always consists of an even number of groups. Since Citizens are randomly placed into groups, this allows us to keep the distribution of Corruptibility Index balanced between the Citizens of each treatment condition. Third, we subtly inform the initial Committee Members of the selection rule to nudge them to act in a corrupt/honest manner. Specifically, initial Committee Members in the Corrupt Committee Treatment see the following text:

> You have the role of Committee Member.
>
> This assignment is based on your decisions in Part 1 of the experiment.
>
> For each of the activities you played in Part 1, we assigned 1 point to participants who did **not** donate, 1 point to those who did **not** invest in the group account, 1 point to those who reported a **large** number of tails and 1 point to those who decided to **take** from the group. We assigned 0 points otherwise.
>
> You and the other two participants have been chosen as Committee Members because you scored the highest in the four activities of Part 1.

In the Honest Committee Treatment, the three Committee Members see a similar text that shows the opposite selection rule. Only the initial Committee Members are informed of the selection rule. Subsequent members only see a brief description of the role.

3.3.4 Sample and Procedure

From Fall 2022 to Spring 2023, we recruited a total of 224 undergraduate students from Texas A&M University to participate in fourteen experimental sessions. Thus, each session consists of 16 subjects, with 8 subjects in each treatment condition. Table 3.2 reports the number of subjects by treatment condition and initial role assignment. Subjects were recruited using ORSEE (Greiner, 2015). Sessions took place at the Experimental Research Laboratory at Texas A&M University. The experiments were programmed using oTree (Chen et al., 2016). Upon arrival, subjects are assigned a payment ID and seated at computer stations. To ensure anonymity, there is always an empty station between two subjects. A general instructions were read aloud by the experimenter. We told subjects that we will not ask for their names at any time during the experimental session and therefore, no one, including the experimenter, will link subjects' names to the decisions they made in the experiment. Since the four activities in the first part of the experiment are shown in random order to each subject, all instructions are shown on the subjects' screen so each subject can participate at their own pace. Once all subjects complete part one, the program pauses and the instructions for part two were read aloud by the experimenter. The instructions were also shown on subjects' screen. A quiz was shown in the next screen to test whether subjects understood the nature of the game and tasks of each roles.

Each session lasted for about 90 minutes. At the conclusion of the last round of the game, subjects are asked to fill out a brief questionnaire for demographic information and are then shown their earnings from the session. Subjects' payment consists of four parts: 1) a show up fee of \$10; 2) one randomly selected activity from part 1; 3) one randomly selected round from part 2; 4) belief elicitation for randomly selected sequence from part 2. Payments are made privately to subjects through either Venmo or Paypal. The average payoff is \$23.79 including the show-up fee.

Table 3.2: Treatment Conditions

Treatment	Initial Role	# Subjects
Honest Committee	Citizen	70
	Committee Member	42
Corrupt Committee	Citizen	70
	Committee Member	42

3.3.5 Estimation Strategy

We test the effects of the status quo level of embezzlement on Citizen's decision to join the committee by estimating the following equation using OLS regression.

$$Selection_{it} = \alpha + \beta_1 CorruptTreatment_i + \delta X_i + \lambda_t + \epsilon_{it} \tag{3.3}$$

$Selection_i$ is a binary variable that equals 1 if the Citizen answered "yes" to the question "Would you like to become a Committee Member?". $CorruptTreatment_i$ is an indicator variable that equals 1 if subject i is assigned to the Corrupt Committee Treatment. The Honest Committee Treatment is the omitted group in regression analysis. X_i is a set of individual characteristics collected from the endline survey, including the subject's age, gender, major, previous experience in economic experiment, and the number of other participants the subject knows in the session. β_1 captures the extent to which corrupt committees with high levels of embezzlement affect Citizens' self-selection. λ_t is the round fixed effect that captures whether self-selection varies by time.

As we show in our model a Citizen's self-selection may be affected by his propensity for corruption. For example, corrupt Citizens might be overall more willing to serve on the committee. It is also possible that our treatment conditions might affect self-selection differently for subjects with different levels of propensity for corruption. For example, corrupt subject might be more willing to join a committee in the Corrupt Committee Treatment and less willing to join a committee in the Honest Committee Treatment. In order to statistically test whether there are heterogeneous treatment effects by subjects' own propensity for corruption, we estimate equation 3.4 below.

$$\begin{aligned}Selection_{it} = \alpha + \beta_1 CorruptTreatment_i + \beta_2 Index_i \\ + \beta_3 CorruptTreatment_i * Index_i + \delta X_i + \lambda_t + \epsilon_{it}\end{aligned} \tag{3.4}$$

In this specification, β_1 captures the treatment effects on the least corrupt (most honest and prosocial) subjects. β_3 captures the differential impacts of treatment on subjects who have a higher propensity for corrupt behaviors.

As described in Section 3.3.1.4, the Corruptibility Index measures a subject's propensity for corruption along two dimensions: honesty vs dishonesty, and prosociality vs self-interest. To see whether these two characteristics play different roles in the self-selection process, we divide the Corruptibility Index into two sub-indices: the Honesty Index (Taking VCM + Coin Toss), and the Prosocial Index (Donation + Giving VCM). We then estimate the following equation:

$$\begin{aligned}Selection_{it} = \alpha + \beta_1 CorruptTreatment_i + \beta_2 HonestyIndex_i \\ + \beta_3 CorruptTreatment_i * HonestyIndex_i \\ + \beta_4 ProsocialIndex_i + \beta_5 CorruptTreatment_i * ProsocialIndex_i \\ + \delta X_i + \lambda_t + \epsilon_{it}\end{aligned} \tag{3.5}$$

In this specification, β_3 captures the differential impacts of treatment on subjects who have a higher propensity for honest behaviors. β_5 captures the differential impacts of treatment on subjects who have a higher propensity for prosocial behaviors.

As a secondary analysis, we examine Committee Members' voting behavior for embezzlement by estimating the following equation.

$$\begin{aligned}Embezzle_{it} = \alpha + \beta_1 CorruptTreatment_i \\ + \beta_2 CorruptTreatment_i * Index_i + \gamma Index_i \\ + \sigma NewMember_{it} + \delta X_i + \lambda_t + \epsilon_{it}\end{aligned} \tag{3.6}$$

The outcome variable $Embezzle_{it}$ is a binary variable that equals 1 if Committee Member i voted for embezzlement in round t. $NewMember_{it}$ is an indicator variable that equals 1 if Committee Member i in round t is the newest member of the committee. Other variables are similarly defined as before. β_1 captures the treatment effects on the least corrupt subjects. β_2 captures the differential impacts of treatment on subjects who have a higher propensity for corrupt behaviors. γ allows us to examine whether the Corruptibility Index is a good predictor for Committee Members' embezzlement behavior. If members with a higher Corruptibility Index are more likely to vote for embezzlement, we should expect γ to be positive and statistically significant. Finally, σ captures whether newly joined Committee Members are more likely to vote for embezzlement.

3.4 Results

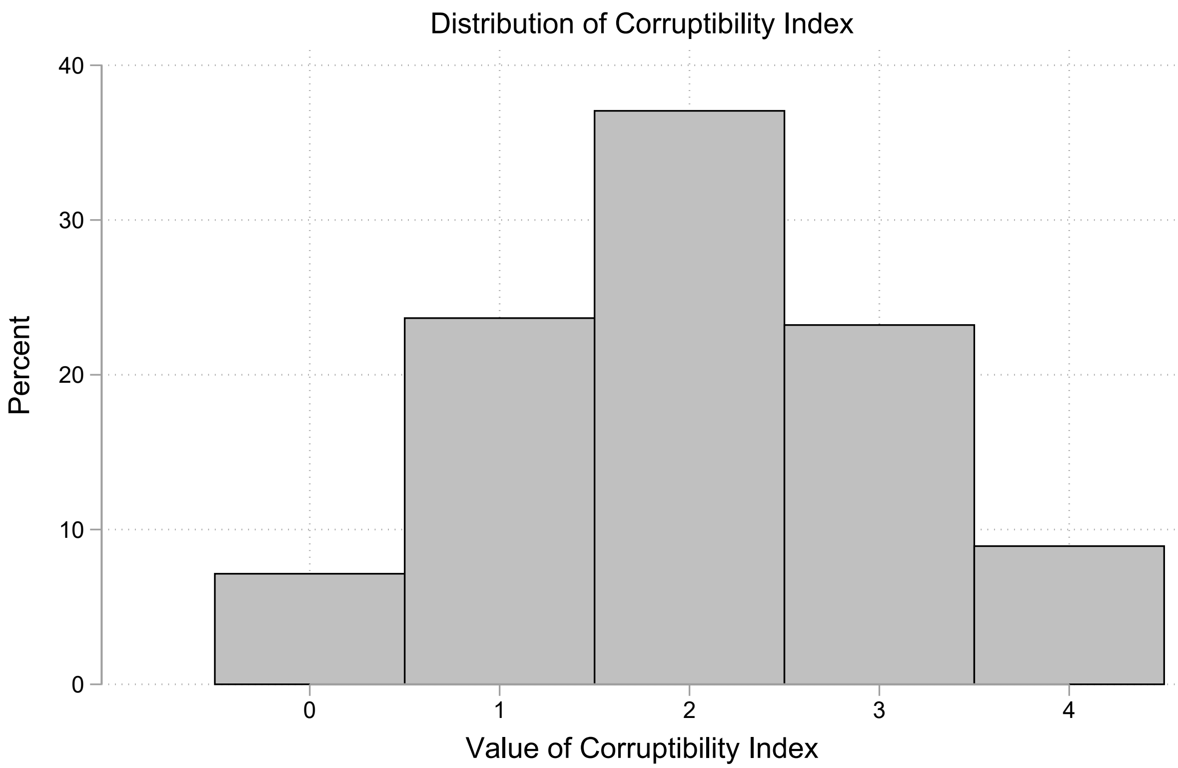
In this section, we present the results from our experiment. Section 3.4.1 reports our findings from the Pre-games. Section 3.4.2 reports the treatment effects on overall embezzlement activities and Citizens' beliefs. Section 3.4.3 reports the results on Citizens' self-selection into the committee. Section 3.4.4 reports the results on Committee Members' voting behavior. Section 3.4.5 reports the results on Committee Members' chat analysis.

3.4.1 The Pre-games

In this subsection, we present findings from the Pre-games. We conducted 4 one-shot social preference games to characterize subjects' overall corruptibility. In the Donation Game, 55% of the subjects kept the money to themselves instead of donating to a charity organization. In the Giving VCM Game, 67% of the subjects chose to invest in the Private Account. In the Taking VCM Game, 64% of the subjects chose to take from the Group Account. In the Coin Toss Game, 15% of the subjects reported more than 11 tails.

As described in Section 3.3.1.4, we assign 1 point for each of the above "undesirable" behaviors and use the sum of points to construct the Corruptibility Index. Figure 3.1 displays the distribution of the index. The majority of the data points are clustered toward the middle. 37.05% of the subjects has a Corruptibility Index of 2. The average value is 2.03. The distribution tapers off toward either extreme. About 7% of the subjects has a corruptibility Index of 0 while 9% of the subject has a corruptibility Index of 4.

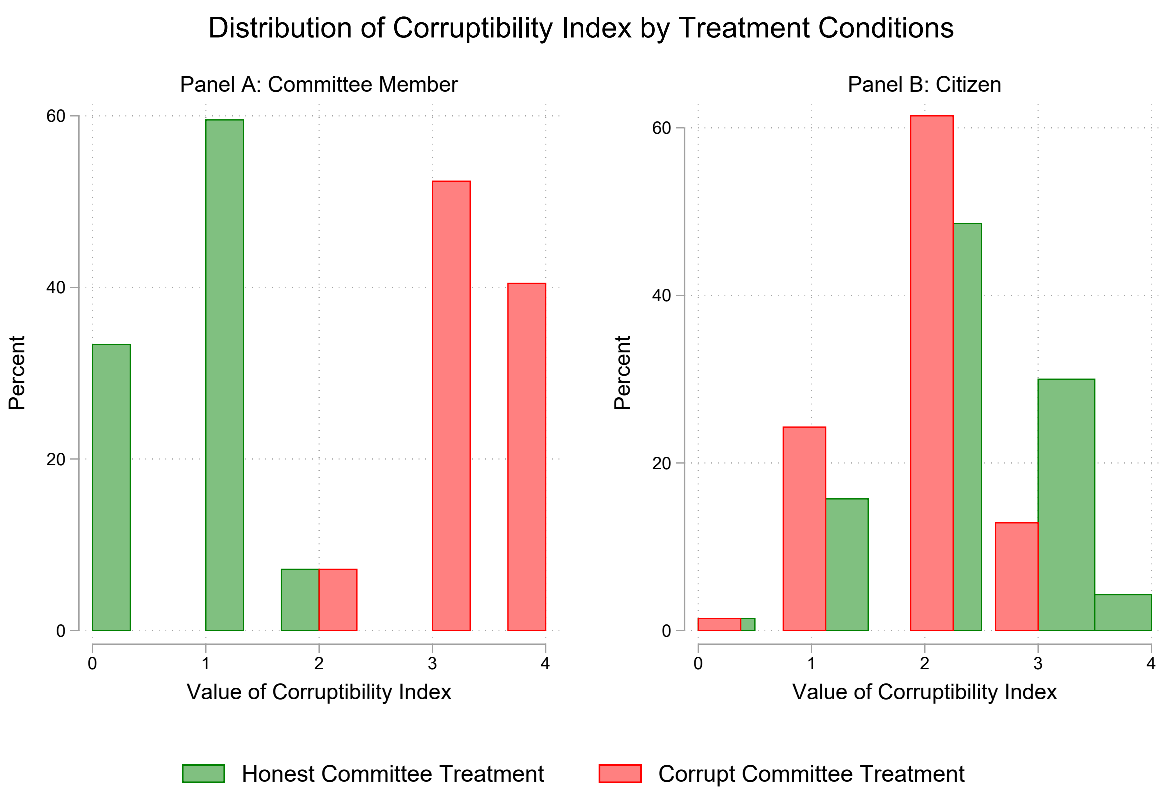
Figure 3.1: Distribution of the Corruptibility Index



Notes: This figure shows the distribution of the Corruptibility Index among all 224 subjects. This index is created using subjects' responses in the Pre-gamesas described in Section 3.3.1.4. The maximum value of the index is 4, indicating the subject being the most corrupt. The minimum value of the index is 0, indicating the subject is the least corrupt.

By design, we exogenously place the three subjects with the highest Corruptibility Index in the session into the initial committee in the Corrupt Committee Treatment. The selection of the initial committee in the Honest Committee Treatment follows the opposite rule. Panel A of Figure 3.2 displays the distribution of the initial Committee Members' Corruptibility Index by treatment conditions. As the result of our treatment manipulation, Committee Members in the Corrupt Committee Treatment have significantly higher Corruptibility Index (3.33 vs 0.74, p-value=0.00). Subjects who are initially assigned the role of the Citizen are randomly shuffled between the Corrupt Committee Treatment and the Honest Committee Treatment to avoid possible imbalances. Panel B of Figure 3.2 displays the distribution of the initial Citizens' Corruptibility Index by treatment conditions. Despite the randomization, Citizens in the Corrupt Committee Treatment are significantly less corrupt compared to Citizens in the Honest Committee Treatment (1.86 vs 2.2, p-value=0.01). This could be caused by an insufficient sample size. We account for this imbalance in our analyses by controlling for the Corruptibility Index.

Figure 3.2: Distribution of the Corruptibility Index by Treatment Conditions



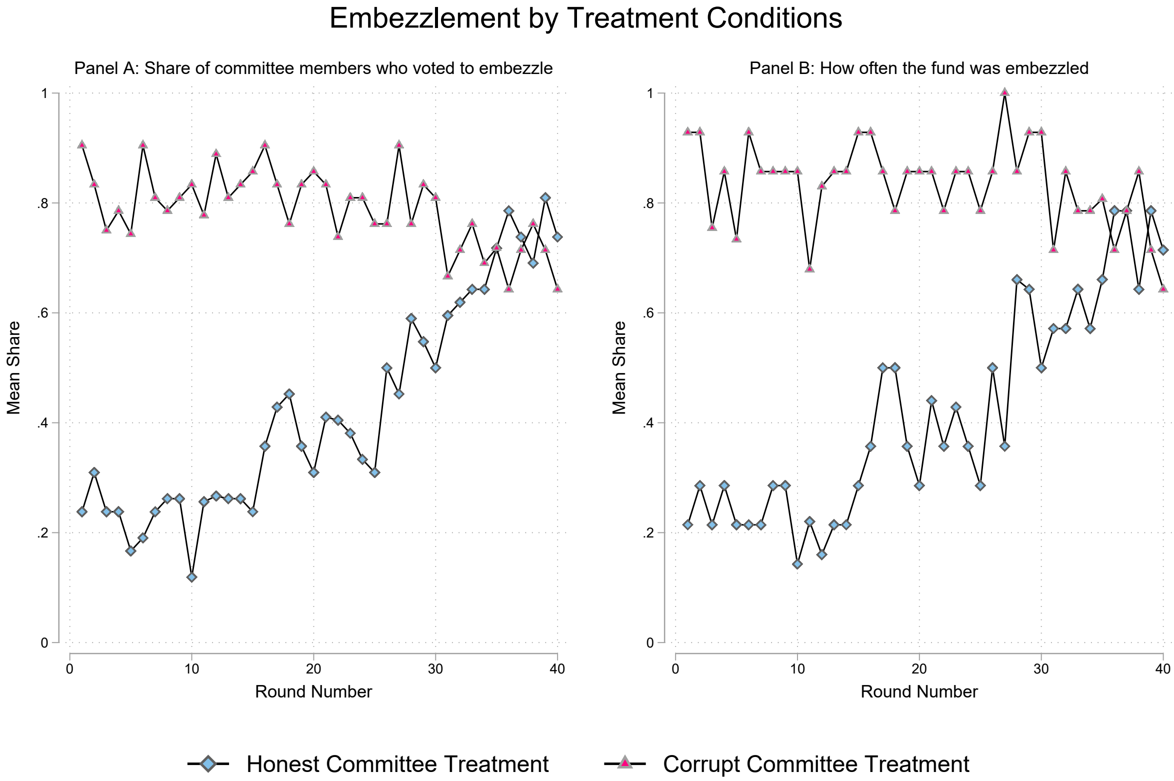
Notes: This figure shows the distribution of the Corruptibility Index by treatment conditions and by initial role assignments. This index is created using subjects' responses in the Pre-gamesas described in Section 3.3.1.4. The maximum value of the index is 4, indicating the subject being the most corrupt. The minimum value of the index is 0, indicating the subject is the least corrupt. Panel A shows the distribution for subjects who are initially assigned the role of Committee Member. Panel B shows the distribution for subjects who are initially assigned the role of Citizens. The Green bar represents Honest Committee Treatment, the Red bar represents Corrupt Committee Treatment.

3.4.2 Embezzlement and Beliefs

In the last subsection, we showed that our treatment conditions generated two contrasting types of committee. Committee Members in the Corrupt Committee Treatment have significantly higher Corruptibility Index than Committee Members in the Honest Committee Treatment. The purpose of this treatment manipulation is to create initial committees with different status quo level of embezzlement. Thus, as a first step, we examine whether embezzlement of the public fund occurs more frequently in the Corrupt Committee Treatment. Panel A of Figure 3.3 plots the share of Committee Members who voted for embezzlement in each round. Panel B of Figure 3.3 plots the frequency of embezzlement in each round. The two figures share similar patterns. Embezzlement occur significantly more frequently in the Corrupt Committee Treatment, although the gap gradually shrinks and is reversed in the last sequence. Thus, we successfully created initial committees that are corrupt in the Corrupt Committee Treatment, and honest in the Honest Committee Treatment.

Next, we examine whether Citizens are able to update beliefs about the Committee's embezzlement level. Since Citizens do not receive perfect information about the Committee's action, for our treatment manipulation to work, Citizens must be able to form beliefs after observing the history of investment outcomes. For every five round, Citizens are asked to guess the number of times the Committee Members kept the money for themselves in the past five round. Figure 3.4 plots Citizens' average beliefs about embezzlement by treatment conditions. For comparison, we also plot the the average number of times that the public fund was lost in Panel A, and average number of times that the Committee's voting resulted in embezzlement in Panel B. Despite the lack of transparency, Citizens' beliefs about embezzlement in the Corrupt Committee Treatment is significantly higher than the Honest Committee Treatment. Moreover, Citizens' beliefs are largely accurate as they closely align with actual embezzlement. Thus, even when Citizens do not perfectly observe the Committee's actions, they are able to infer (nearly) correctly about that the Committee is corrupt. We successfully generated societies with different levels of embezzlement and showed that Citizens update their beliefs in response to the embezzlement.

Figure 3.3: Embezzlement by Treatment Conditions

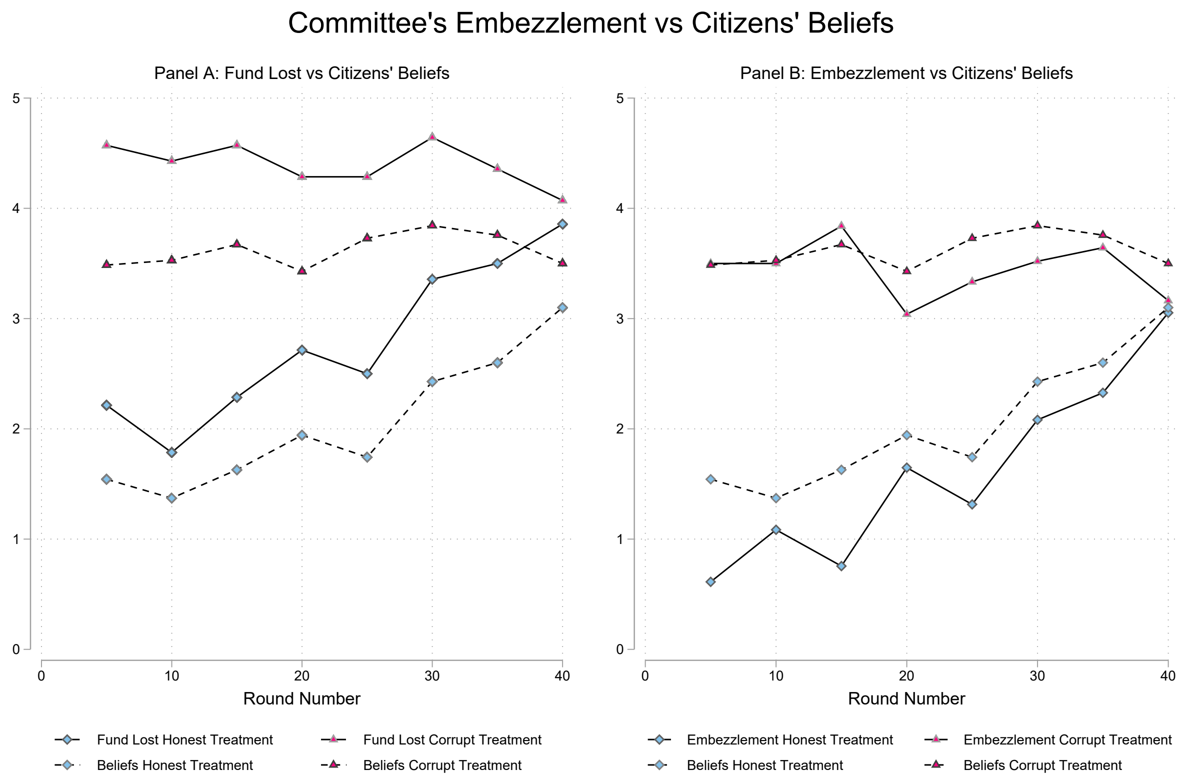


Notes: This figure plots the occurrence of embezzlement each round by treatment conditions. Panel A shows the share of Committee Members who voted for embezzlement in each round. Panel B shows the frequency of embezzlement in each round. The blue diamond line represents the Honest Committee Treatment, the pink triangle line represents the Corrupt Committee Treatment.

3.4.3 Self-selection

We now turn to the analysis of our main research question, who self-select into committees. In the Committee Game, Citizens have the chance to periodically self-select into the committee at the beginning of every new sequence. Figure 3.5 plots the share of Citizens who expressed interest in joining the committee. First, committees in the Corrupt Committee Treatment attract more Citizens than committees in the Honest Committee Treatment. This is true for every sequence of the game. Second, Citizens' willingness to serve in the committee shows a slight decline over time in both treatment conditions. We formalize the graphical evidence in Figure 3.5 using regressions. Column 1 reports estimates from equation 3.3. The dependent variable is a binary variable that

Figure 3.4: Embezzlement vs Beliefs by Treatment Conditions

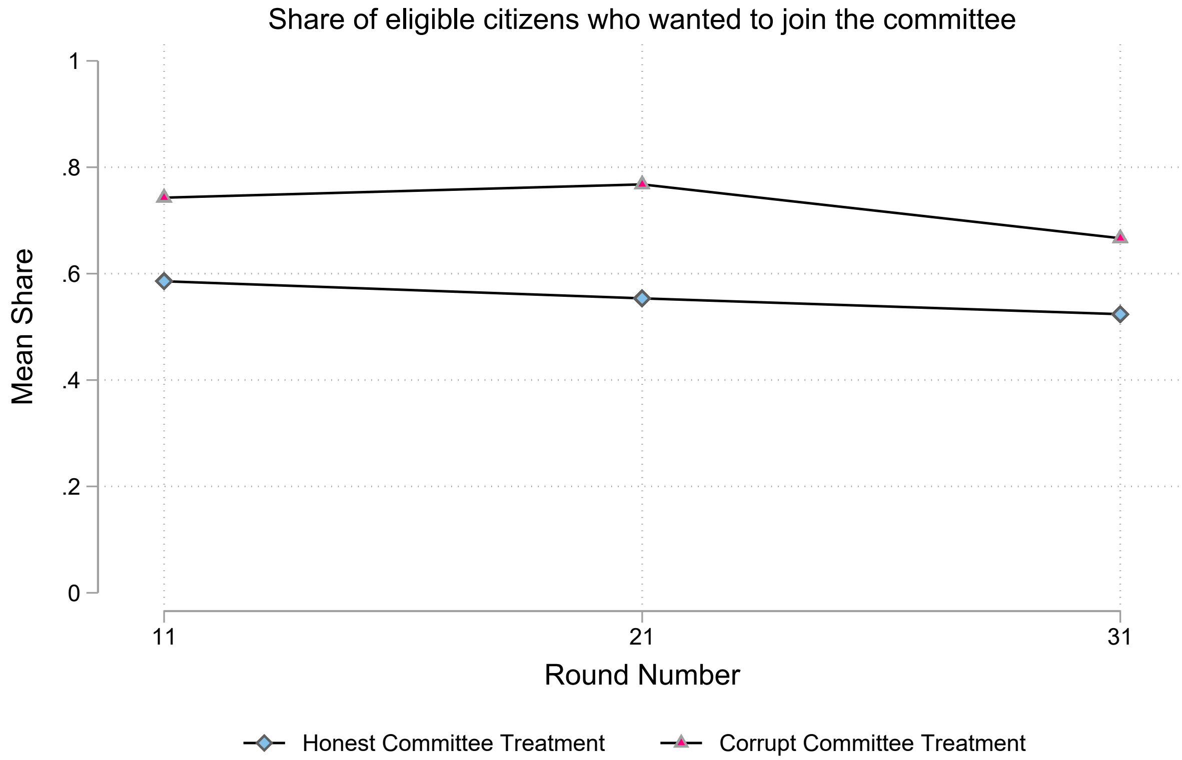


Notes: This figure shows the comparison between Citizens' beliefs about embezzlement and Committees' embezzlement. Panel A shows the average number of times that the public fund was lost, which could be determined by either the Committee's choice to embezzle, or by luck. Panel B shows the average number of times that the Committee's voting resulted in embezzlement, which does not involve luck. The blue diamond line represents the Honest Committee Treatment. The pink triangle line represents the Corrupt Committee Treatment.

equals 1 if the Citizen answered "yes" to the question "Would you like to become a Committee Member?". The regression estimates confirm the gap in self-selection between treatment conditions. Citizens in the Corrupt Committee Treatment are 17.3 percentage points more likely to want to join the committee. Column 2 additionally controls for the Corruptibility Index. The coefficient on the treatment dummy is similar in size as the one in Column 1. However, the coefficient on the Corruptibility Index is close to 0 and does not show statistical significance.

While the estimates in Column 2 show that subjects' propensity for corruption has little effect on their interest in joining the committee, it is possible that different subjects are attracted to

Figure 3.5: Share of Eligible Citizens who Wanted to Join the Committee



Notes: This figure plots the share of eligible citizens who answer "yes" to the question "Would you like to become a Committee Member?" in the beginning of each decision sequence. The blue diamond line represents the Honest Committee Treatment. The pink triangle line represents the Corrupt Committee Treatment.

different committees. As our model predicts, corrupt Citizens are more likely to join corrupt committees and honest Citizens are more likely to join honest committees. This heterogeneity could cause the overall lack of effects. We examine this possibility by estimating equation 3.4. This specification additionally controls for the interaction term between the treatment dummy and the Corruptibility Index. We report the estimates in Column 3. The coefficient on the treatment dummy loses significance. Moreover, the coefficient on the interaction term between the treatment dummy and the Corruptibility Index is close to zero and is not statistically significant. Thus, we do not find heterogeneous treatment effect by individual propensity for corruption. Our findings suggest that Citizens are more inclined to join the Committee when they have experienced embezzlement

Table 3.3: The Decision to Join the Committee

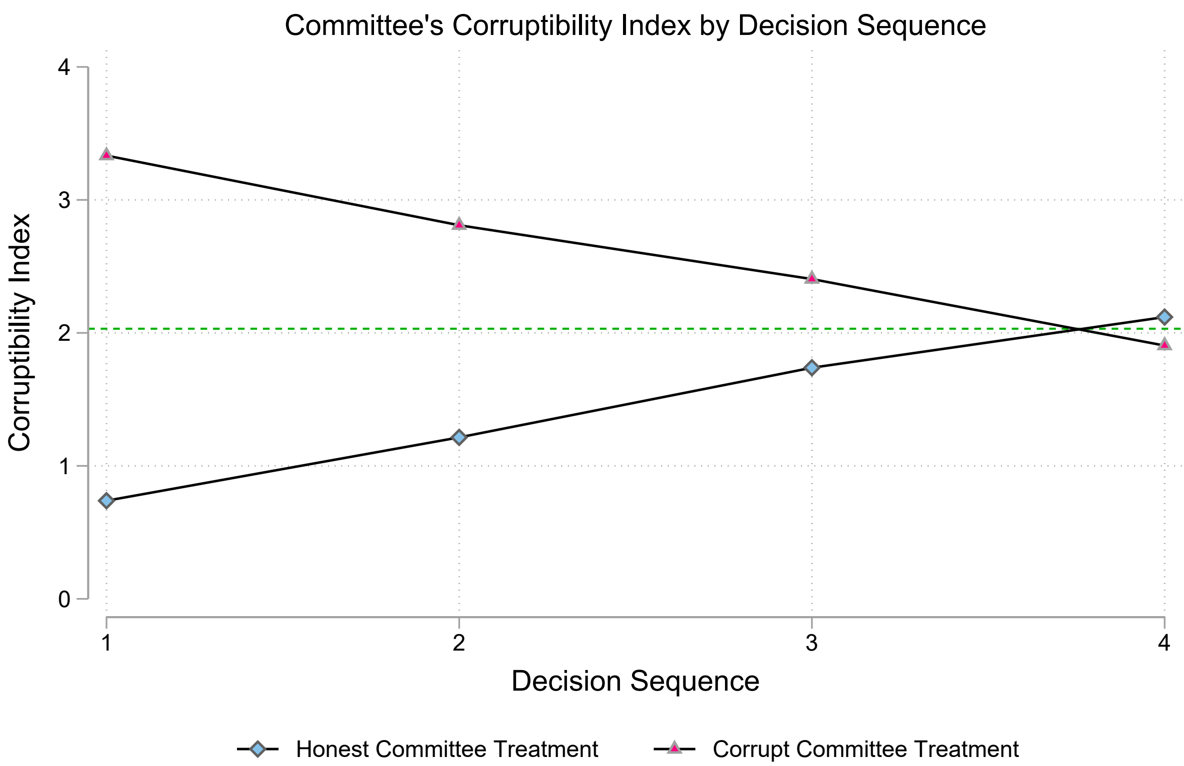
	Wants to join the Committee (0-1)		
	(1)	(2)	(3)
Corrupt Treatment	0.173**	0.168*	0.309
	(0.078)	(0.082)	(0.237)
Corruptibility Index		-0.014	0.015
		(0.056)	(0.083)
C Treatment * Corruptibility Index			-0.071
			(0.109)
Round Number	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)
Observations	336	336	336
R-squared	0.035	0.036	0.038
Control Mean	0.560	0.560	0.560

Notes: This table presents results from OLS regressions. The dependent variable is an indicator variable that equals 1 if the Citizen answered "yes" to the question "Would you like to become a Committee Member?". The independent variables include a dummy for the Corrupt Committee Treatment, the Corruptibility Index, the interaction term between the treatment dummy and the index, and round number. Standard errors in parentheses and clustered at the society level. *** p<0.01, ** p<0.05, * p<0.1

themselves. This pattern is true for corrupt Citizens, as well as honest and prosocial Citizens.

The null effect of the Corruptibility Index on self-selection is also evident in Figure 3.6. Due to our treatment manipulation, at the start of the game, the Committees in each treatment condition have very contrasting values of Corruptibility Index. However, as the game progresses, the average value of Corruptibility Index increases in Corrupt Committees decreases and decreases in Honest Committees. By the last decision sequence, the two lines intersect at around $y = 2$, which happens to be the mean Corruptibility Index of the entire subject pool. The convergence to the mean provides further evidence that the Corruptibility Index does not predict self-selection into the Committee. Furthermore, it explains the convergence of embezzlement activity shown in Figure 3.3. Compared to the initial Committees, as the Committees in the Corrupt (Honest) Committee Treatment become more honest (corrupt), the frequency of embezzlement naturally goes down (up).

Figure 3.6: Committee's Corruptibility Index by Treatment Conditions



Notes: This figure plots the average value of Corruptibility Index of Committees in each decision sequence. The blue diamond line represents the Honest Committee Treatment. The pink triangle line represents the Corrupt Committee Treatment. The green dotted line plots $y = 2.03$, the mean Corruptibility Index of the entire subject pool.

Next, as specified in Section 3.3.5, we decompose the Corruptibility Index into two components, the Honesty Index and the Prosocial Index. Although estimates reported in Table 3.3 suggest that Citizens are more likely to select into Committees in the Corrupt Committee Treatment regardless of their own "types", it is possible that the two indices may play different roles in the self-selection process, and this may be masked when they are combined into one aggregate index. Thus, we examine this possibility by estimating equation 3.5. The estimates reported in Table 3.4 are largely in line with the estimates in Table 3.3. While the coefficient on the treatment dummy remains large and significant in all specification except for Column 2, neither sub-indices nor the interaction term shows significance. Thus, we do not find heterogeneous treatment effect by the individual's honesty or prosociality.

Table 3.4: The Decision to Join the Committee

	Wants to join the Committee (0-1)			
	(1)	(2)	(3)	(4)
Corrupt Treatment	0.168**	-0.019	0.194**	0.269**
	(0.078)	(0.138)	(0.079)	(0.107)
Honesty Index	0.080	0.005		
	(0.054)	(0.085)		
C Treatment * Honesty Index		0.150		
		(0.103)		
Prosocial Index			-0.072	-0.030
			(0.055)	(0.086)
C Treatment * Prosocial Index				-0.100
				(0.099)
Round Number	-0.003	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	336	336	336	336
R-squared	0.045	0.054	0.042	0.045
Control Mean	0.560	0.560	0.560	0.560

Notes: This table presents results from OLS regressions. The dependent variable is an indicator variable that equals 1 if the Citizen answered "yes" to the question "Would you like to become a Committee Member?". The independent variables include a dummy for the Corrupt Committee Treatment, the Honesty Index, the Prosocial Index, the interaction term between the treatment dummy and each index, and round number. Standard errors in parentheses and clustered at the society level. *** p<0.01, ** p<0.05, * p<0.1

3.4.4 Voting

Our findings in the last subsection indicate that Citizens are more likely to self-select into Corrupt Committees. This pattern is true regardless of the Citizen's propensity for corruption, i.e., whether the Citizen is corrupt or honest. Thus, the decision to self-select does not seem to depend on the Citizen's type. However, it is possible that different types of Citizen choose to join the Committee for different reasons. For example, a corrupt Citizen might join because he wants to maximize his monetary gains through embezzlement, an honest and prosocial Citizen might join because he wants to stop the embezzlement by past members. To examine why Citizens join the Committee, we examine their voting behavior once they become a Committee Member.

Table 3.5: The Decision to Embezzle

	Vote for embezzlement (0-1)			
	(1)	(2)	(3)	(4)
Corrupt Treatment	0.367***	0.180*	0.272	0.234
	(0.102)	(0.097)	(0.173)	(0.182)
Corruptibility Index		0.162***	0.182***	0.174***
		(0.020)	(0.036)	(0.038)
C Treatment * Corruptibility Index			-0.044	-0.026
			(0.057)	(0.063)
New Member				0.053*
				(0.031)
Decision Sequence	0.058**	0.058***	0.047*	0.046*
	(0.025)	(0.018)	(0.026)	(0.025)
Observations	3,312	3,312	3,312	3,312
R-squared	0.159	0.265	0.266	0.268
Control Mean	0.422	0.422	0.422	0.422

Notes: This table presents results from OLS regressions. The dependent variable is an indicator variable that equals 1 if the Committee Member voted for embezzlement. The independent variables include a dummy for the Corrupt Committee Treatment, the Corruptibility Index, the interaction term between the treatment dummy and the index, a dummy for new members, and dummies for decision sequence. Standard errors in parentheses and clustered at the society level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.5 reports estimates from equation 3.6. Column 1 confirms the graphical evidence in Figure 3.3. Committee Members in the Corrupt Committee Treatment are significantly more likely to vote for embezzlement by 36.7 percentage points. We also find a positive coefficient on the block dummies. This suggests that Committee Members become more likely to vote for embezzlement as the game progresses. However, once we additionally control for the Corruptibility Index in Column 2, the coefficient on the treatment dummy becomes nearly halved, although still statistically significant. The Corruptibility Index strongly predicts voting behavior. Every additional increase in the value of the Committee Member's Corruptibility Index would significantly increase his likelihood of voting for embezzlement by 16.2 percentage points. To investigate possible heterogeneous treatment effects, Column 3 additionally controls for the interaction term between the treatment dummy and the Corruptibility Index. However, similar to our previous analyses, the coefficient on the interaction term is very small and does not show statistical significant. Thus, we do not find evidence that the Corruptibility Index predicts voting differently in our two treatment conditions. Column 4 additionally controls for a dummy for new members who just joined the Committee in that block. Our estimates indicate that new Committee Members are more likely to vote for embezzlement by 4.6 percentage points. This could be cause by two reasons. Either new members are persuaded by existing members to vote for embezzlement, or they joined with the intention to embezzle.

Our analyses in Table 3.5 show that Committee Members vote in line with their type. The more (less) corruptible types are more (less) likely to vote for embezzlement, regardless of other factors such as the status quo level of embezzlement. This voting pattern seems to suggest that corrupt individuals are entering to take advantage of embezzlement, while the honest and prosocial individuals are entering to either keep the Committee "clean", or "clean up" the corrupt Committee. However, to reveal Citizens' motivation behind self-selection, more analysis is needed. Given the structure of the game, a subject can serve on the Committee for as short as 10 rounds, or as long as 30 rounds. Thus, the data structure is imbalanced as some individuals have more observations. The longer the subject stays in the Committee, the more likely his decision will be affected by his fellow Committee Members. Thus, only the decisions made during the first 10 rounds the subject serve on the Committee can best reflect the subject's intention of self-selection. For easier interpretation, we further divide the full sample into Honest Committee Treatment and Corrupt Committee Treatment. Finally, we separate the Corruptibility Index into the Honesty Index and the Prosocial Index as before to examine whether different traits play different roles. We report the results in Table 3.6. Column 1 to Column 3 reveals a consistent pattern: the Prosocial Index is a strong predictor for voting behavior for new Committee Members in the Honest Committee Treatment. The more prosocial the subject is, the more likely that he will vote against embezzlement in the Honest Committee Treatment. In contrast, Column 4 to Column 6 consistently shows that the Honesty Index is a strong predictor for voting behavior for new Committee Members in the Corrupt Committee Treatment. The more honest the subject is, the more likely that he will vote against embezzlement in the Corrupt Committee Treatment. These findings suggest that the honesty trait and the prosocial trait play different roles in different treatment conditions. However, since we know from previous results in Table 3.4 that neither traits play a role in the self-selection process, it must be the case that these traits are activated after the subject joined the Committee. We investigate this possibility in the next subsection through the analysis of chat messages posted by the Committee Members.

3.4.5 Communication between Committee Members

In this subsection, we analyze the communication between Committee Members to further explore the motivation behind their self-selection. Recall that Committee Members are able to talk with each other via chat for 2 minutes at the beginning of round 1 and round 6 of every decision sequence. We did not explicitly provide instructions on what the should discuss. Thus, Committee Members can discuss anything they like. Table 3.7 provides an overview of the chat messages. Every message is manually categorized into one of the three categories based on its content: 1) proposing the Committee should **embezzle** the money; 2) proposing that the Committee should **share** the money equally with the Citizens; 3) general chit-chat. About 75% of the chat messages involve general chit-chat that does not relate to the experiment. Examples of the Chit-chat category include general greetings such as "how are you", or expressions such as "LOL" or "Haha". A comparison between treatment conditions shows that the communication between Committee Members in the Corrupt Committee Treatment is significantly more likely to feature Embezzling messages (18.77% vs 11.17%, p-value=0.00), and less likely to feature Sharing messages (5.92%

Table 3.6: The Decision to Embezzle and the Decomposition of the Corruptibility Index

	Vote for embezzlement (0-1)					
	Honest Committee			Corrupt Committee		
	(1)	(2)	(3)	(4)	(5)	(6)
Honesty Index	-0.088	0.011	-0.088	-0.123***	-0.116***	-0.123***
	(0.074)	(0.073)	(0.074)	(0.033)	(0.037)	(0.034)
Honesty Index * Round		-0.018**			-0.001	
		(0.007)			(0.004)	
Prosocial Index	-0.252***	-0.251***	-0.192***	-0.039	-0.039	-0.045
	(0.034)	(0.034)	(0.046)	(0.052)	(0.052)	(0.066)
Prosocial Index * Round			-0.011			0.001
			(0.006)			(0.010)
Round	0.008	0.036**	0.020	-0.004	-0.003	-0.004
	(0.008)	(0.014)	(0.012)	(0.004)	(0.006)	(0.007)
Observations	812	812	812	826	826	826
R-squared	0.173	0.176	0.175	0.052	0.052	0.052
Control Mean	0.403	0.403	0.403	0.791	0.791	0.791

Notes: This table presents results from OLS regressions. The dependent variable is an indicator variable that equals 1 if the Committee Member voted for embezzlement. The sample only includes the first 10 rounds the subject serves on the Committee. The left panel shows the results for the Honest Committee Treatment. The right panel shows the results for the Corrupt Committee Treatment. Independent variables include the Honesty Index, the Prosocial Index, round number, and two interaction terms between the indices and the round number. Standard errors in parentheses and clustered at the society level. *** p<0.01, ** p<0.05, * p<0.1

vs 14.6%, p-value-0.00).

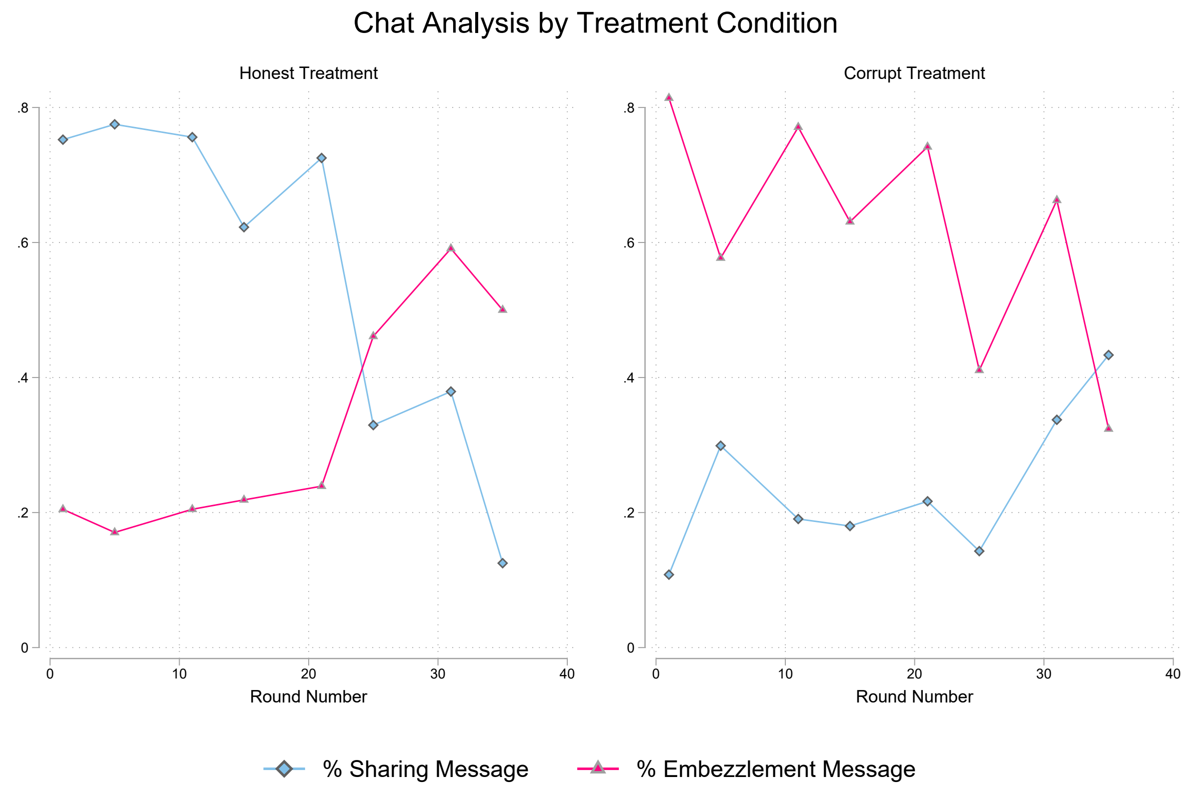
Table 3.7: Overview of Chat

	Full Sample	**Honest Treatment**	**Corrupt Treatment**
# Embezzling	436 (14.67%)	179 (11.17%)	257 (18.77%)
# Sharing	315 (10.6%)	234 (14.6%)	81 (5.92%)
# Chit-chat	2,221 (74.73%)	1,190 (74.24%)	1,031 (75.31%)

Notes: A message is labeled as one of the following category: 1) proposing embezzlement; 2) proposing equal distribution; 3) general chat-chat.

One concern about using chat analysis is that the message posted by Committee Members

Figure 3.7: Committee Member's Message Posted by Treatment Conditions



Notes: This figure plots the types of message posted in each communication round in each treatment condition. The left panel shows the messaging in the Honest Committee Treatment. The right panel shows the messaging in the Corrupt Committee Treatment. The Blue line represents the percentage of message that is related to equally sharing the money with citizens. The Pink line represents the percentage of message that is related to keeping the money only for the Committee Members themselves.

does not necessarily reflect their true intention. For example, a subject could be lying. While this is possible, the anonymous feature of our setting should help reduce this possibility.[18] As a validity check, Figure 3.7 plots the change of chat content overtime. In the Honest Committee Treatment, there is initially very high percentage of Sharing messages and very low percentage of Embezzlement messages. However as the game progresses, the percentage for sharing decreases while the percentage for embezzling increases. Eventually the two line intersects. The opposite dynamic is observed in the Corrupt Committee Treatment. This pattern closely resembles Figure

[18]Each Committee Members is assigned a number which allows Committee Members to associate a message with the member posting it. For example, the screen will show "Committee Member 1: Hi guys." However, the number dose not allow Committee Members to identify each other.

3.3 and shows that the communication between Committee Members is closely aligned with the Committee's decisions. Thus, the content of messages is a valid reflection of subjects' action and thought process.

Table 3.8 shows results from regression analyses. Subjects' communication pattern is largely consistent with the voting behavior. First, Committee Members in the Corrupt Committee Treatment are significantly more likely to send embezzlement messages and significantly less likely to send sharing messages. Second, the Corruptibility Index is a very strong predictor for the types of message a subject sends. Estimates in Column 2 suggest that every additional increase in the value of the Corruptibility Index would increase the subject's likelihood of sending an embezzlement message by 7 percentage points. Similarly, estimates in Column 5 suggest that every additional increase in the value of the Corruptibility Index would decrease the subject's likelihood of sending a sharing message by 3.7 percentage points. Third, we do not find heterogeneous treatment effects by the individual's propensity for corruption. The interaction term between the treatment dummy and the Corruptibility Index is small and insignificant in both Column 3 and Column 6. Thus, just like the voting behavior, the content of Committee Members' communication is also in line with their "types."

In the last subsection, we showed that individuals vote in line with their types. Moreover, the Prosocial Index predicts voting behaviors in the Honest Committee Treatment, while the Honesty Index predicts voting behaviors in the Corrupt Committee Treatment. One possible explanation for this pattern is that the communication between Committee Members in the Honest Committee Treatment made the prosocial trait more salient, while the communication between Committee Members in the Corrupt Committee Treatment made the honesty trait more salient. We test this hypothesis by analysing the chat messages posted by Committee Members. As an illustrative example, Appendix C.2 provides examples of the reasoning that is frequently used by Committee Members during the communication round. The different types of reasoning we see does provide some support for our hypothesis. In particular, Committee Members in the Honest Committee Treatment often justify their support for equal distribution by emphasizing that it is better for

Table 3.8: Chat Analysis

	Embezzlement message			Sharing message		
	(1)	(2)	(3)	(4)	(5)	(6)
Corrupt Treatment	0.076**	-0.013	0.019	-0.088***	-0.037	-0.028
	(0.031)	(0.036)	(0.052)	(0.023)	(0.027)	(0.049)
Corruptibility Index		0.070***	0.077***		-0.037***	-0.035**
		(0.012)	(0.021)		(0.009)	(0.015)
C Treatment * Corruptibility Index			-0.015			-0.005
			(0.028)			(0.019)
Round	-0.000	-0.001	-0.001	-0.003**	-0.002***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.118***	0.034	0.031	0.198***	0.243***	0.242***
	(0.033)	(0.029)	(0.029)	(0.023)	(0.027)	(0.029)
Observations	2,973	2,893	2,893	2,973	2,893	2,893
R-squared	0.012	0.050	0.051	0.030	0.044	0.044

Notes: This table presents results from OLS regressions. The dependent variable in Column 1 to Column 3 is an indicator variable that equals 1 if the message is about embezzlement. The dependent variable in Column 4 to Column 6 is an indicator variable that equals 1 if the message is about sharing. The independent variables include a dummy for the Corrupt Committee Treatment, the Corruptibility Index, the interaction term between the treatment dummy and the index, and round number. Standard errors in parentheses and clustered at the society level. *** p<0.01, ** p<0.05, * p<0.1

the society as a whole. This type of reasoning clearly speaks to the prosocial aspect of equal distribution and is likely to gain support from other Committee Members with high Prosocial Index. In contrast, the prosocial trait will likely play a less important role in the Corrupt Committee Treatment since the communication hardly touches on this aspect. We formally investigate this hypothesis using regression analyses and report the estimates in Table 3.9. Column 1 shows that the Honesty Index and the Prosocial Index play roughly an equal role in predicting the messaging content in the Honest Committee Treatment. Column 2 shows that only the Honesty Index predicts the messaging content in the Corrupt Committee Treatment. Thus, our estimates provide partial support for the hypothesis that the honesty and prosocial traits play different roles in different treatment.

3.5 Conclusion

In this paper, we use a laboratory experiment to examine the self-selection into committees that have discretionary power over the distribution of public resources. We are primarily interested

Table 3.9: Chat Analysis and the Decomposition of the Corruptibility Index

	Embezzlement message	
	Honest Treatment	Corrupt Treatment
	(1)	(2)
Honesty Index	-0.076**	-0.076***
	(0.035)	(0.018)
Prosocial Index	-0.068**	-0.017
	(0.029)	(0.039)
Round	-0.000	-0.003
	(0.002)	(0.002)
Constant	0.309***	0.324***
	(0.101)	(0.041)
Observations	1,575	1,318
R-squared	0.051	0.034

Notes: This table presents results from OLS regressions. The dependent variable is an indicator variable that equals 1 if the message is about embezzlement. The left panel shows the results for the Honest Committee Treatment. The right panel shows the results for the Corrupt Committee Treatment. The independent variables include the Honesty Index, the Prosocial Index, and round number. Standard errors in parentheses and clustered at the society level. *** p<0.01, ** p<0.05, * p<0.1

in how the status quo level of corruption and the individual's propensity for corruption affect the decision to join the committee. Our experiment proceeds in two stages. In the first stage, we run four one-shot social preference games to construct an overall measure of corruptibility for each subject. In the second stage, we run a committee game that simulates a society where Citizens can periodically self-select into a committee. Citizens engage in real effort tasks and contribute to a public fund. Committee Members are in charge of managing the public fund and can choose to steal from the fund without the Citizens knowing. Our treatment conditions manipulate the composition of the initial committee to examine how the status quo level of corruption affect self-selection.

Our first finding is that our treatment conditions successfully created two contrasting type of committees. Committees in the Corrupt Committee Treatment are significantly more likely to embezzle from the public fund compared to committees in the Honest Committee Treatment. Our second finding is that Citizens are able to form accurate beliefs about the committee's embezzlement activity, although Citizens only observe the amount they receive from the public fund. Our third finding is that both corrupt and honest Citizens are more likely to select into committees when they have experienced past corruption. Our fourth finding is that Committee Members vote in line with their type regardless of the status quo level of embezzlement. In both treatment conditions, more corrupt types are more likely to vote for embezzle, and less corrupt types are more likely to vote for fair redistribution. Our fifth finding is the honesty trait and the prosocial trait play different role in determining Committee Members' decision, this seems to be activated by the communication between Committee Members.

Our paper provides new evidence on the self-selection into public service positions. Our experimental design allows us to examine both the role of individual characteristics and the status quo level of corruption in one framework. Contrary to Hanna and Wang (2017) and its replication studies, we find that individual propensity for corruption does not predict the selection. However, we find suggestive evidence that the propensity for corruption does predict the embezzlement behavior of subjects who selected into committees. Our results highlight the importance of screening for public service positions such as government employees. We show that subjects' behavior in a series of one-shot social preference games are highly correlated with embezzlement behavior. Given that these short games are quick and easy to implement, it could potentially be a cost-effective way of reducing corruption. Our results have implications for future work. A natural follow up question to our study is, how can we induce more prosocial types to join the committee? Since people vote according to types, then having more prosocial types in the committee should decrease the likelihood of corruption. As next steps, we plan to take advantage of the chat data between Committee Members to conduct a more detailed analysis of dynamics of play within the Committee. For example,are corruptible players trying to convince other members to be corrupt? We also plan to test whether mechanisms that resemble town hall meetings and require Committee Members to communicate their decisions to the public affect both corruption decision-making and self-selection into committees.

4. SUMMARY AND CONCLUSIONS

My dissertation studies topics on hate crime, media and corruption. In the first chapter, titled "Does News Coverage of Hate-motivated Mass Shootings Generate More Hatred?" I investigate the role news media plays in promoting hatred through the news coverage of mass shootings. While there has been a long-standing debate on how media should report mass shootings, my paper is the first to causally test whether news coverage of mass shootings can increase support for the shooter and the shooter's hate-driven ideology. I first show through observational data that whenever a mass shooting is targeting a specific race/ethnicity/gender/sexual orientation, it receives higher media coverage with more focus on the shooter. Using online searching data and hate crime data, I show that this difference in coverage is correlated with viewers' reactions and possibly results in an increase in hatred toward the victimized group in the shooting. Based on these findings and the existing literature, I design and conduct an online information provision experiment that manipulates how a mass shooting targeting immigrants is reported in the news. The study involves more than 2,000 individuals living in the United States. I stratify the randomization by political affiliation, to be able to test whether and how the information reported in the news has different effects depending on subjects' ex-ante beliefs about immigration, which are captured by Democrat or Republican party affiliation. I show that details on the shooter's ideology increases Republican subjects' support for the shooter. Emphasizing the shooter's identity and background, while highlighting past victimization and possible mental health problems, increases Democrat subjects' support for both the shooter and the shooter's anti-immigration ideology. I also find suggestive evidence that this exposure increases the interest of Democrat subjects' in white supremacy hate groups. Further analysis shows that the treatment effects on Democrats are driven by the more right-lining individuals within the sample. Overall, my study highlights the unintended consequences of news coverage of hate-motivated mass shootings. My findings provide important guidelines to the media's approach to reporting on such shootings and, more broadly, on crimes that have the potential to impact individual attitudes toward both the hate ideology of the suspect and the victimized group.

In the second chapter, titled "Can Social Media Rhetoric Incite Hate Incidents? Evidence from Trump's "Chinese Virus" Tweets" and co-authored with Jason Lindo and fellow PhD student Jiee Zhong, I focus on social media and anti-Asian incidents. In particular, my coauthors and I investigate whether former president Donald Trump's tweets, in which he referred to COVID-19 as the "Chinese Virus," contributed to the rise of anti-Asian incidents. Although many papers have shown high-profile individuals can promote pro-social behaviors like interest in preventative health care and voting, there is little evidence whether this kind of influence can extend to anti-social behaviors. We use an event-study framework and show that the number of incidents spiked following Trump's initial "Chinese Virus" tweets and there was a subsequent dramatic rise in internet search activity for the phrase. Moreover, the spike in anti-Asian incidents was significantly more pronounced in counties that supported Donald Trump in the 2016 presidential election relative to those that supported Hillary Clinton. Overall, this study shows that high-profile individuals such as Trump can have detrimental effects, even when the technology of social media substantially limits what they can say. Our findings have important implications given the recent rise of populist leaders pushing antisocial beliefs and behaviors on topics ranging from vaccine hesitancy to the treatment of immigrants. This paper received a Revise & Resubmit request from the Journal of Urban Economics, and we recently completed the revision and resubmitted it.

The third chapter of my dissertation is titled "Who Self-selects into Committees: the Pro-social or the Corrupt?" and is joint with Dmitry Ryvkin (Florida State University) and my advisor Danila Serra. This study focuses on another one of my research interests, i.e., individual corruptibility in group settings. We build a theoretical model of a citizen's decision to join a committee that has discretion over the distribution of a public fund generated by taxes levied on the work/income of citizens. Committee members – who earn less than citizens – decide whether to distribute the fund equally among members of the society, or to embezzle the money. The model predicts that citizens' beliefs play a major role. An increase in one's own corruptibility or corruptibility of the committee—increases willingness to join for corrupt citizens but decreases willingness to join for honest citizens. We employ a laboratory experiment and a university student sample to test these predictions. Our results show that subjects are more likely to select into committees when they have experienced past corruption. This is true regardless of subjects' own propensity for corruption. Moreover, we find that Committee Members' decision on whether to embezzle is in line with their "types" regardless of the status quo level of embezzlement. More corrupt types are more likely to support embezzlement, and less corrupt types are more likely to support fair redistribution. Finally, we show that the honesty trait and the prosocial trait play different role in determining Committee Members' decision, this seems to be activated by the communication between Committee Members. This study provides new evidence on the self-selection into public service positions. Our results highlight the importance of screening for public service positions such as government employees, along characteristics such honesty and prosociality.

My dissertation is a reflection of my research interests in crime, corruption, and other anti-social behaviors. By employing both experimental and quasi-experimental methodologies, I aim at producing research findings that are not only causal, but also valid both in and out of experimental settings to address policy-relevant questions. In the years to come, I plan to continue working on topics related to hate, crime and media. First, I aim to expand my research on media coverage of mass shootings by exploring different variants of shootings and news stories. One question that I was not able to address in the first chapter of this dissertation is whether the media could shift the attention from the suspects in hate-motivated shootings to the victims and survivors, possibly leading to less hatred. An immediate and important extension of my work would be to test whether news coverage of victims' identities and backgrounds could improve individuals' attitudes and behaviors toward the victimized group. I also plan to extend the third chapter of this dissertation. A natural follow up question to our study is, how can we induce more honest and prosocial types to enter the committee? While the existing literature largely focuses on monetary punishment, given the novelty of our design and results, we plan to vary the accountability to the public and examine how it changes the Committee's action and the Citizens' self-selection. We plan to conduct a Town Hall Meeting treatment that simulates bottom-up accountability through communication to and pressure from citizens. Another research topic I plan to pursue in the near future relates to the growing political polarization in the United States, and the perceptions of political correctness and cancel culture. In particular, I am interested in empirically investigating the impact of the cancel culture movement on individuals' beliefs and behaviors towards others who hold similar or different views, and the motives behind individuals' participation in a "cancelling" movement. As part of my preliminary motivational analysis, I researched two high-profile cancel culture cases against celebrities: JK Rowling and Joe Rogan. I collected the universe of tweets that mentions their name during the course of their cancellation. I show that while the majority of tweets are calling for accountability, a large proportion of tweets contains harassment and insulting content. In the near future, I plan to design an experiment to understand why people may actively engage in canceling attempts, with an emphasis on the desire to hold wrongdoers accountable versus the desire to signal one's type to relevant peers to avoid backlash.

REFERENCES

Abbink, K., Irlenbusch, B., & Renner, E. (2002). An experimental bribery game. *Journal of Law, economics, and organization*, *18*(2), 428–454.

Abeler, J., Becker, A., & Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, *113*, 96–104.

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics*, *130*(4), 1885–1939.

Alatas, V., Chandrasekhar, A. G., Mobius, M., Olken, B. A., & Paladines, C. (2019). *When celebrities speak: A nationwide twitter experiment promoting vaccination in indonesia* (Tech. Rep.). National Bureau of Economic Research.

Alesina, A., Miano, A., & Stantcheva, S. (2018). *Immigration and redistribution* (Tech. Rep.). National Bureau of Economic Research.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, *103*(484), 1481–1495.

Andreoni, J. (1995). Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, *110*(1), 1–21.

Ang, D. (2020). The birth of a nation: Media and racial hate.

Arias, E. (2019). How does media influence social norms? experimental evidence on the role of common knowledge. *Political Science Research and Methods*, *7*(3), 561–578.

Ashraf, N., Bandiera, O., Davenport, E., & Lee, S. S. (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review*, *110*(5), 1355–94.

Attanasi, G., Rimbaud, C., & Villeval, M. C. (2019). Embezzlement and guilt aversion. *Journal of Economic Behavior & Organization*, *167*, 409–429.

Austin, E. W., Vord, R. V. d., Pinkleton, B. E., & Epstein, E. (2008). Celebrity endorsements and their potential to motivate young voters. *Mass communication and society*, *11*(4), 420–436.

Azfar, O., & Nelson Jr, W. R. (2007). Transparency, wages, and the separation of powers: An experimental analysis of corruption. *Public Choice*, 471–493.

Banerjee, R., Baul, T., & Rosenblat, T. (2015). On self selection of the corrupt into the public sector. *Economics Letters*, *127*, 43–46.

Banuri, S., & Eckel, C. (2015). Cracking down on bribery. *Social Choice and Welfare*, *45*, 579–600.

Banuri, S., & Keefer, P. (2016). Pro-social motivation, effort and the call to public service. *European Economic Review*, *83*, 139–164.

Barfort, S., Harmon, N. A., Hjorth, F., & Olsen, A. L. (2019). Sustaining honesty in public service: The role of selection. *American Economic Journal: Economic Policy*, *11*(4), 96–123.

Barr, A., Lindelow, M., & Serneels, P. (2009). Corruption in public service delivery: An experimental analysis. *Journal of Economic Behavior & Organization*, *72*(1), 225–239.

Barr, A., Packard, T., & Serra, D. (2014). Participatory accountability and collective action: Experimental evidence from albania. *European Economic Review*, *68*, 250–269.

Barr, A., & Serra, D. (2010). Corruption and culture: An experimental analysis. *Journal of Public economics*, *94*(11-12), 862–869.

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological inquiry*, *7*(1), 1–15.

Bénabou, R., Falk, A., & Tirole, J. (2018). *Narratives, imperatives, and moral reasoning* (Tech. Rep.). National Bureau of Economic Research.

Bénabou, R., & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, *126*(2), 805–855.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507.

Bertrand, M., Burgess, R., Chawla, A., & Xu, G. (2020). The glittering prizes: Career incentives and bureaucrat performance. *The Review of Economic Studies*, *87*(2), 626–655.

Bertrand, M., Djankov, S., Hanna, R., & Mullainathan, S. (2007). Obtaining a driver's license in india: an experimental approach to studying corruption. *The Quarterly Journal of Economics*, *122*(4), 1639–1676.

Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American economic review*, *95*(3), 616–636.

Besley, T., & Ghatak, M. (2018). Prosocial motivation and incentives. *Annual Review of Economics*, *10*, 411–438.

Brassiolo, P., Estrada, R., Fajardo, G., & Vargas, J. (2021). Self-selection into corruption: Evidence from the lab. *Journal of Economic Behavior & Organization*, *192*, 799–812.

Bursztyn, L., Egorov, G., Enikolopov, R., & Petrova, M. (2019). *Social media and xenophobia: evidence from russia* (Tech. Rep.). National Bureau of Economic Research.

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

Bursztyn, L., Haaland, I. K., Rao, A., & Roth, C. P. (2020). *Disguising prejudice: Popular rationales as excuses for intolerant expression* (Tech. Rep.). National Bureau of Economic Research.

Bursztyn, L., Rao, A., Roth, C. P., & Yanagizawa-Drott, D. H. (2020). *Misinformation during a pandemic* (Tech. Rep.). National Bureau of Economic Research.

Bushman, B. J. (2018). Narcissism, fame seeking, and mass shootings. *American behavioral scientist*, *62*(2), 229–241.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Cameron, L., Chaudhuri, A., Erkal, N., & Gangadharan, L. (2009). Propensities to engage in and punish corrupt behavior: Experimental evidence from australia, india, indonesia and singapore. *Journal of public economics*, *93*(7-8), 843–851.

Cao, A., Lindo, J. M., & Zhong, J. (2022). *Can social media rhetoric incite hate incidents? evidence from trump's" chinese virus" tweets* (Tech. Rep.). National Bureau of Economic Research.

Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., … Jurafsky, D. (2022). Computational analysis of 140 years of us political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, *119*(31), e2120510119.

Card, D., & Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The quarterly journal of economics*, *126*(1), 103–143.

Carpenter, J., & Seki, E. (2011). Do social preferences increase productivity? field experimental evidence from fishermen in toyama bay. *Economic Inquiry*, *49*(2), 612–630.

Carr, J., Clifton-Sprigg, J., James, J., & Vujic, S. (2020). *Love thy neighbour? brexit and hate crime* (Tech. Rep.). IZA Discussion Papers.

Cartwright, E. (2016). A comment on framing effects in linear public good games. *Journal of the Economic Science Association*, *2*(1), 73–84.

Center, P. R. (2019). For local news, americans embrace digital but still want strong community connection. *Pew Research Center*.

Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74–87.

Charness, G., & Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives*, *26*(3), 157–176.

Chen, D. L., Schonger, M., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Chopra, F., Haaland, I., & Roth, C. (2022). Do people demand fact-checked news? evidence from us democrats. *Journal of Public Economics*, *205*, 104549.

Chou, H.-Y. (2015). Celebrity political endorsement effects: A perspective on the social distance of political parties. *International Journal of Communication*, *9*, 24.

Cohn, A., & Maréchal, M. A. (2018). Laboratory measure of cheating predicts school misconduct. *The Economic Journal*, *128*(615), 2743–2754.

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies*, *82*(4), 1289–1308.

Couttenier, M., Hatte, S., Thoenig, M., & Vlachos, S. (2021). Anti-muslim voting and media coverage of immigrant crimes. *The Review of Economics and Statistics*, 1–33.

Cox, C. A., & Stoddard, B. (2015). Framing and feedback in social dilemmas with partners and strangers. *Games*, *6*(4), 394–412.

Cram, P., Fendrick, A. M., Inadomi, J., Cowen, M. E., Carpenter, D., & Vijan, S. (2003). The impact of a celebrity promotional campaign on the use of colon cancer screening: the katie couric effect. *Archives of Internal Medicine*, *163*(13), 1601–1605.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, *24*(4), 349.

Dai, Z., Galeotti, F., & Villeval, M. C. (2018). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, *64*(3), 1081–1100.

Dal Bó, E., Finan, F., & Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, *128*(3), 1169–1218.

DeFoster, R., & Swalve, N. (2018). Guns, culture or mental health? framing mass shootings as a public health crisis. *Health communication*, *33*(10), 1211–1222.

DellaVigna, S., & Gentzkow, M. (2010). Persuasion: empirical evidence. *Annu. Rev. Econ.*, *2*(1), 643–669.

DellaVigna, S., & Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, *122*(3), 1187–1234.

Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., & Jurafsky, D. (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

De Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266–3302.

Deserranno, E. (2019). Financial incentives as signals: experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics*, *11*(1), 277–317.

Dhar, D., Jain, T., & Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. *American Economic Review*, *112*(3), 899–927.

Di Falco, S., Magdalou, B., Masclet, D., Villeval, M. C., & Willinger, M. (2016). Can transparency of information reduce embezzlement? experimental evidence from tanzania. *Experimental Evidence from Tanzania (May 17, 2016). UMR*, *5824*.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, *73*, 66–88.

Doherty, C., Kiley, J., & Asheer, N. (2019). In a politically polarized era, sharp divides in both partisan coalitions. *PEW Research*.

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., Sears, R. R., Ford, C. S., ... Sollenberger, R. T. (2013). *Frustration and aggression*. Routledge.

Durante, R., & Zhuravskaya, E. (2018). Attack when the world is not watching? us news and the israeli-palestinian conflict. *Journal of Political Economy*, *126*(3), 1085–1133.

Duxbury, S. W., Frizzell, L. C., & Lindsay, S. L. (2018). Mental illness, the media, and the moral politics of mass violence: The role of race in mass shootings coverage. *Journal of Research in Crime and Delinquency*, *55*(6), 766–797.

Eisensee, T., & Strömberg, D. (2007). News droughts, news floods, and us disaster relief. *The Quarterly Journal of Economics*, *122*(2), 693–728.

Ekins, E. E. (2017). The state of free speech and tolerance in america; attitudes about free speech, campus speech, religious liberty, and tolerance of political expression. *Findings From the Cato Institute*.

Engel, C. (2011). Dictator games: A meta study. *Experimental economics*, *14*, 583–610.

Evans, D. G. R., Barwell, J., Eccles, D. M., Collins, A., Izatt, L., Jacobs, C., ... others (2014). The angelina jolie effect: how high celebrity profile can have a major impact on provision of cancer related services. *Breast Cancer Research*, *16*(5), 1–6.

Falk, A., & Fischbacher, U. (2002). "crime" in the lab-detecting social interaction. *European Economic Review*, *46*(4-5), 859–869.

Fehr, E., & Leibbrandt, A. (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of public economics*, *95*(9-10), 1144–1155.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, *20*(2), 303–315.

Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. Cambridge University Press.

Fox, J. A., & Fridel, E. E. (2016). The tenuous connections involving mass shootings, mental illness, and gun laws. *Violence and Gender*, *3*(1), 14–19.

Friebel, G., Kosfeld, M., & Thielmann, G. (2019). Trust the police? self-selection of motivated agents into the german police force. *American Economic Journal: Microeconomics*, *11*(4), 59–78.

Fujiwara, T., Müller, K., & Schwarz, C. (2021). *The effect of social media on elections: Evidence from the united states* (Tech. Rep.). National Bureau of Economic Research.

Gans-Morse, J., Kalgin, A., Klimenko, A., Vorobyev, D., & Yakovlev, A. (2021). Self-selection into public service when corruption is widespread: The anomalous russian case. *Comparative Political Studies*, *54*(6), 1086–1128.

Garthwaite, C., & Moore, T. J. (2013). Can celebrity endorsements affect political outcomes? evidence from the 2008 us democratic presidential primary. *The journal of law, economics, & organization*, *29*(2), 355–384.

Gerber, A. S., Karlan, D., & Bergan, D. (2009). Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, *1*(2), 35–52.

Ginges, J., & Atran, S. (2009). What motivates participation in violent political action: selective incentives or parochial altruism? *Annals of the New York Academy of Sciences*, *1167*(1), 115–

123.

Gino, F., Gu, J., & Zhong, C.-B. (2009). Contagion or restitution? when bad apples can motivate ethical behavior. *Journal of Experimental Social Psychology*, *45*(6), 1299–1302.

Gneezy, U., Saccardo, S., & Van Veldhuizen, R. (2019). Bribery: Behavioral drivers of distorted decisions. *Journal of the European Economic Association*, *17*(3), 917–946.

Gover, A. R., Harper, S. B., & Langton, L. (2020). Anti-asian hate crime during the covid-19 pandemic: Exploring the reproduction of inequality. *American journal of criminal justice*, *45*, 647–667.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, *1*(1), 114–125.

Grigorieff, A., Roth, C., & Ubfal, D. (2020). Does information change attitudes toward immigrants? *Demography*, *57*(3), 1117–1143.

Gupta, N., Rigotti, L., & Wilson, A. (2021). The experimenters' dilemma: Inferential preferences over populations. *arXiv preprint arXiv:2107.05064*.

Haaland, I., & Roth, C. (2020). Labor market concerns and support for immigration. *Journal of Public Economics*, *191*, 104256.

Haaland, I., & Roth, C. (2021). Beliefs about racial discrimination and support for pro-black policies. *The Review of Economics and Statistics*, 1–38.

Haaland, I., Roth, C., & Wohlfart, J. (2020). Designing information provision experiments.

Hanna, R., & Wang, S.-Y. (2017). Dishonesty and selection into public service: Evidence from india. *American Economic Journal: Economic Policy*, *9*(3), 262–90.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1–14).

He, B., Ziems, C., Soni, S., Ramakrishnan, N., Yang, D., & Kumar, S. (2021). Racism is a virus: Anti-asian hate and counterspeech in social media during the covid-19 crisis. In *Proceedings of the 2021 ieee/acm international conference on advances in social networks analysis and mining* (pp. 90–94).

Helfgott, J. B. (2015). Criminal behavior and the copycat effect: Literature review and theoretical framework for empirical investigation. *Aggression and violent behavior*, *22*, 46–64.

Horowitz, J. M., Parker, K., & Stepler, R. (2017). Wide partisan gaps in us over how far the country has come on gender equality. *Pew Research Center*, *18*.

Hswen, Y., Xu, X., Hing, A., Hawkins, J. B., Brownstein, J. S., & Gee, G. C. (2021). Association of "# covid19" versus "# chinesevirus" with anti-asian sentiments on twitter: March 9–23, 2020. *American Journal of Public Health*, *111*(5), 956–964.

Jackson, D. J., & Darrow, T. I. (2005). The influence of celebrity endorsements on young adults' political opinions. *Harvard international journal of press/politics*, *10*(3), 80–98.

Jetter, M. (2017). The effect of media attention on terrorism. *Journal of Public Economics*, *153*, 32–48.

Jetter, M., & Walker, J. K. (2022). News coverage and mass shootings in the us. *European Economic Review*, 104221.

Khadjavi, M., & Lange, A. (2015). Doing good or doing harm: experimental evidence on giving and taking in public good games. *Experimental Economics*, *18*, 432–441.

Kissner, J. (2016). Are active shootings temporally contagious? an empirical assessment. *Journal of Police and Criminal Psychology*, *31*(1), 48–58.

Klein, J. (2012). Bully society, the. In *Bully society, the.* New York University Press.

Konstabel, K., Lönnqvist, J.-E., Walkowitz, G., Konstabel, K., & Verkasalo, M. (2012). The 'short five'(s5): Measuring personality traits using comprehensive single items. *European Journal of Personality*, *26*(1), 13–29.

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

Langman, P. (2018). Different types of role model influence and fame seeking among mass killers and copycat offenders. *American Behavioral Scientist*, *62*(2), 210–228.

Lankford, A. (2016). Fame-seeking rampage shooters: Initial findings and empirical predictions.

Aggression and violent behavior, *27*, 122–129.

Lankford, A., & Cowan, R. G. (2020). Has the role of mental health problems in mass shootings been significantly underestimated? *Journal of Threat Assessment and Management*, *7*(3-4), 135.

Lankford, A., & Madfis, E. (2018a). Don't name them, don't show them, but report everything else: A pragmatic proposal for denying mass killers the attention they seek and deterring future offenders. *American behavioral scientist*, *62*(2), 260–279.

Lankford, A., & Madfis, E. (2018b). Media coverage of mass killers: Content, consequences, and solutions. *American Behavioral Scientist*, *62*(2), 151–162.

Lantz, B. (2022). Women who commit hate-motivated violence: Advancing a gendered understanding of hate crime. *Social science research*, *104*, 102682.

Lee, K. J. (2018). Mass shootings and media contagion theory: Social media's influence on frequency of incidents. *Elon Journal of Undergraduate Research in Communications*, *9*(2), 27–35.

Levin, B. (2021). Report to the nation: Anti-asian prejudice and hate crime.

Lim, C. S., Snyder Jr, J. M., & Strömberg, D. (2015). The judge, the politician, and the press: newspaper coverage and criminal sentencing across electoral systems. *American Economic Journal: Applied Economics*, *7*(4), 103–35.

Martin, G. J., & Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, *107*(9), 2565–99.

Mastrorocco, N., & Minale, L. (2018). News media and crime perceptions: Evidence from a natural experiment. *Journal of Public Economics*, *165*, 230–255.

McGinty, E. E., Kennedy-Hendricks, A., Choksy, S., & Barry, C. L. (2016). Trends in news media coverage of mental illness in the united states: 1995–2014. *Health Affairs*, *35*(6), 1121–1129.

McGinty, E. E., Webster, D. W., Jarlenski, M., & Barry, C. L. (2014). News media framing of serious mental illness and gun violence in the united states, 1997-2012. *American journal of public health*, *104*(3), 406–413.

Meindl, J. N., & Ivy, J. W. (2017). Mass shootings: The role of the media in promoting generalized imitation. *American journal of public health*, *107*(3), 368–370.

Metzl, J. M., & MacLeish, K. T. (2015). Mental illness, mass shootings, and the politics of american firearms. *American journal of public health*, *105*(2), 240–249.

Müller, K., & Schwarz, C. (2020). From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103*.

Mummolo, J., & Peterson, E. (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review*, *113*(2), 517–529.

Murray, J. L. (2017). Mass media reporting and enabling of mass shootings. *Cultural Studies Critical Methodologies*, *17*(2), 114–124.

Newman, B., Merolla, J. L., Shah, S., Lemi, D. C., Collingwood, L., & Ramakrishnan, S. K. (2021). The trump effect: An experimental investigation of the emboldening effect of racially inflammatory elite communication. *British Journal of Political Science*, *51*(3), 1138–1159.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in indonesia. *Journal of political Economy*, *115*(2), 200–249.

Olken, B. A., & Pande, R. (2012). Corruption in developing countries. *Annu. Rev. Econ.*, *4*(1), 479–509.

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

Ortiz-Ospina, E., & Roser, M. (2016). Corruption. *Our World in Data*. (https://ourworldindata.org/corruption)

Park, E.-S. (2000). Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding. *Journal of Economic Behavior & Organization*, *43*(4), 405–421.

Parks, J., Bechtold, D., Shelp, F., Lieberman, J., & Coffey, S. (2019). Mass violence in america: Causes, impacts and solutions.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662.

Pezzella, F. S., Fetzer, M. D., & Keller, T. (2019). The dark figure of hate crime underreporting. *American Behavioral Scientist*, 0002764218823844.

Philippe, A., & Ouss, A. (2018). “no hatred or malice, fear or affection”: media and sentencing. *Journal of Political Economy*, *126*(5), 2134–2178.

Reinikka, R., & Svensson, J. (2005). Fighting corruption to improve schooling: Evidence from a newspaper campaign in uganda. *Journal of the European economic association*, *3*(2-3), 259–267.

Reny, T. T., & Barreto, M. A. (2022). Xenophobia in the time of pandemic: Othering, anti-asian attitudes, and covid-19. *Politics, Groups, and Identities*, *10*(2), 209–232.

Roberts, M. C., & Dusetzina, S. B. (2017). The effect of a celebrity health disclosure on demand for health care: trends in brca testing and subsequent health services use. *Journal of community genetics*, *8*(2), 141–146.

Rose-Ackerman, S. (1975). The economics of corruption. *Journal of public economics*, *4*(2), 187–203.

Saad, L. (2013). Americans fault mental health system most for gun violence. *Gallup Politics*, *20*.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annu. Rev. Econ.*, *6*(1), 253–272.

Sacerdote, B., Sehgal, R., & Cook, M. (2020). *Why is all covid-19 news bad news?* (Tech. Rep.). National Bureau of Economic Research.

Salmon, T. C., & Serra, D. (2017). Corruption, social judgment and culture: An experiment. *Journal of Economic Behavior & Organization*, *142*, 64–78.

Schneider-Strawczynski, S., & Valette, J. (2021). Media coverage of immigration and the polarization of attitudes.

Serra, D., Serneels, P., & Barr, A. (2011). Intrinsic motivations and the non-profit health sector: Evidence from ethiopia. *Personality and Individual Differences*, *51*(3), 309–314.

Settele, S. (2022). How do beliefs about the gender wage gap affect the demand for public policy? *American Economic Journal: Economic Policy*, *14*(2), 475–508.

Shleifer, A., & Vishny, R. W. (1993). Corruption. *The quarterly journal of economics*, *108*(3), 599–617.

Silva, J. R., & Greene-Colozzi, E. A. (2019). Fame-seeking mass shooters in america: Severity, characteristics, and media coverage. *Aggression and Violent Behavior*, *48*, 24–35.

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

Song, L. (2022). *Essays on information technology and inequality* (Unpublished doctoral dissertation). New York University.

Sundell, A. (2014). Are formal civil service examinations the most meritocratic way to recruit civil servants? not in all countries. *Public Administration*, *92*(2), 440–457.

Swanson, J. W., McGinty, E. E., Fazel, S., & Mays, V. M. (2015). Mental illness and reduction of gun violence and suicide: bringing epidemiologic research to policy. *Annals of epidemiology*, *25*(5), 366–376.

Wang, T. (2021). Media, pulpit, and populist persuasion: evidence from father coughlin. *American Economic Review*, *111*(9), 3064–92.

Wojcik, S., Hughes, A., & Remy, E. (2019). About one-in-five adult twitter users in the us follow trump.

Xiong, H. (2021). The political premium of television celebrity. *American Economic Journal: Applied Economics*, *13*(4), 1–33.

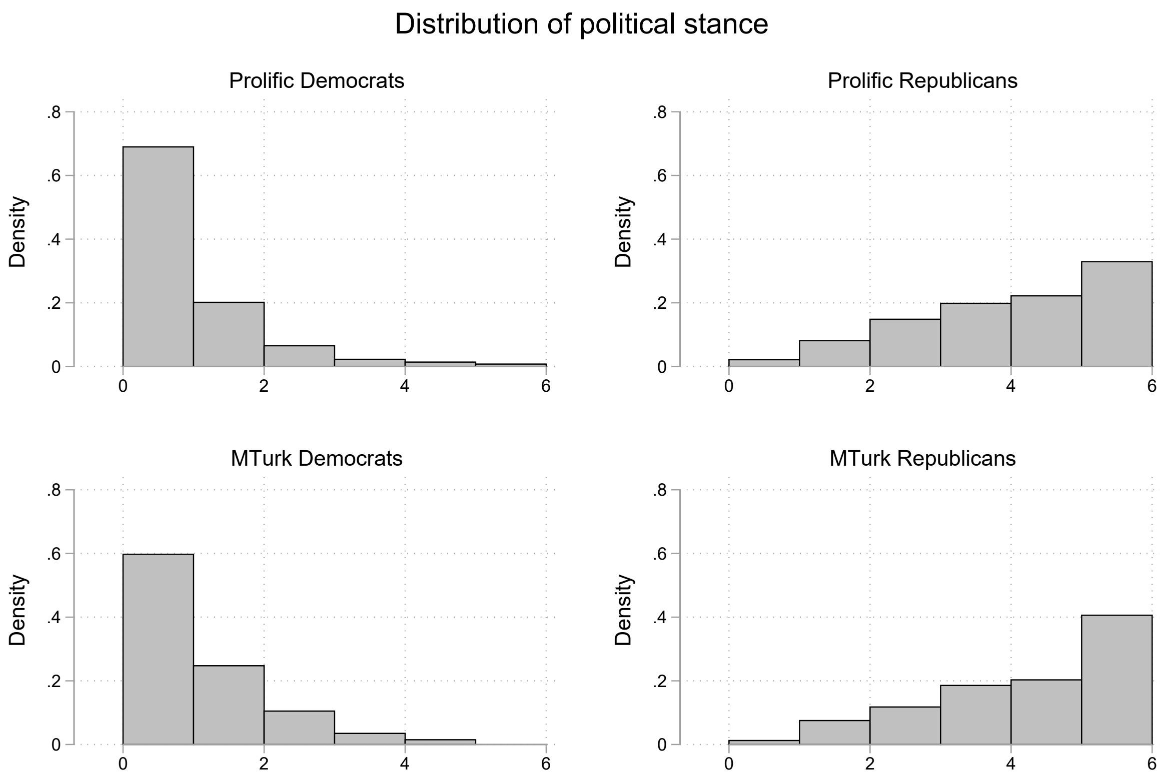
Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the rwandan genocide. *The Quarterly Journal of Economics*, *129*(4), 1947–1994.

APPENDIX A

CHAPTER I APPENDIX

A.1 Appendix A: Additional Tables and Figures

Figure A.1: Distribution of Political Stance



Notes: This graph shows the distribution of political stance. I elicit subjects' opinions on 6 political issues including abortion, same-sex marriage, gun control, minimum wage, build the wall, and citizenship for children of illegal immigrants. An answer that aligns with the Democratic Party ideology will be coded as a 0, an answer that aligns with the Republican Party ideology will be coded as a 1. The political stance is calculated as the summation of all responses.

Table A.1: List of Hate-motivated Mass Shootings

Date	Location	Dead	Type
March 16, 2021	Atlanta, Georgia	8	Anti-Asian
December 10, 2019	Jersey City, New Jersey	6	Anti-Jew
December 6, 2019	Pensacola, Florida	4	Islamic terrorism
August 3, 2019	El Paso, Texas	23	Anti-Hispanic
April 27, 2019	Poway, California	1	Anti-Jew
November 2, 2018	Tallahassee, Florida	3	Anti-female
October 27, 2018	Pittsburgh, Pennsylvania	11	Anti-Jew
September 24, 2017	Antioch, Tennessee	1	Anti-White
April 13, 2017	Fresno, California	4	Anti-White
July 17, 2016	Baton Rouge, Louisiana	4	Anti-White
July 7, 2016	Dallas, Texas	6	Anti-White
June 12, 2016	Orlando, Florida	50	Anti-Gay
December 2, 2015	San Bernardino, California	16	Islamic terrorism
August 26, 2015	Moneta, Virginia	3	Anti-White
July 16, 2015	Chattanooga, Tennessee	6	Islamic terrorism
June 17, 2015	Charleston, South Carolina	9	Anti-Black
May 23, 2014	Isla Vista, California	7	Anti-female
August 5, 2012	Oak Creek, Wisconsin	7	Anti-Sikh
November 5, 2009	Fort Hood, Texas	14	Islamic terrorism
July 28, 2006	Seattle, Washington	1	Anti-Jew
August 10, 1999	Los Angeles, California	1	Anti-Jew
July 2, 1999	Illinois and Indiana	3	Anti-Jew, Anti-Black, Anti-Asian
November 3, 1979	Greensboro, North Carolina	5	Anti-Black
December 31, 1972	New Orleans, Louisiana	10	Anti-White

Table A.2: Summary Statistics of Shootings

	Hate-motivated	**Non-hate-motivated**
# Dead	8.46	6.20
# Injured	10.29	7.84
# Observations	24	217

Notes: A hate-motivated mass shooting is defined as a mass shooting which is motivated, in whole or in part, by the offender's bias(es) against a: race, religion, sexual orientation, ethnicity, gender, gender identity.

Table A.3: Total Minutes per Day on Evening TV News Broadcasts

	Total minutes of news coverage per day			
	(1)	(2)	(3)	(4)
Hate	11.066***	10.259***	6.192***	6.456***
	(0.979)	(0.935)	(1.222)	(1.311)
#Dead		0.323***	0.181***	0.207***
		(0.050)	(0.057)	(0.069)
#Injured		0.067***	0.080***	0.084***
		(0.013)	(0.013)	(0.014)
#Dead * Hate			0.511***	0.515***
			(0.100)	(0.111)
Constant	2.904***	0.295	1.101**	2.343
	(0.321)	(0.412)	(0.439)	(3.330)
Observations	1,561	1,561	1,561	1,561
R-squared	0.076	0.165	0.179	0.297
Fixed Effects				YES

Notes: Column 1 shows an OLS regression of the total minutes of news coverage per day on a dummy for hate-motivated mass shootings. Column 2 additionally controls for the number of dead victims and the number of injured victims. Column 3 adds an interaction term between the number of dead victims and hate. Column 4 adds a set of fixed effects, including days since shooting fixed effect (from day 0, i.e., the day the shooting happened, to day 6, i.e., the 6th day since the shooting happened), day of the week fixed effect, month fixed effect, year fixed effect. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Percentage of news articles about the shooter

	Percentage of News Articles about the Shooter							
	Based on Keywords				Based on Title			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hate	0.043**	0.046**	0.068***	0.050**	0.033**	0.032**	0.052**	0.050**
	(0.018)	(0.018)	(0.024)	(0.025)	(0.015)	(0.016)	(0.021)	(0.022)
#Dead		-0.001	0.001	-0.001		0.000	0.001	-0.000
		(0.001)	(0.001)	(0.002)		(0.001)	(0.001)	(0.001)
#Injured		0.000	-0.000	0.000		-0.000	-0.000	0.000
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
#Dead * Hate			-0.003	-0.002			-0.002	-0.002
			(0.002)	(0.002)			(0.002)	(0.002)
Constant	0.235***	0.239***	0.232***	0.436***	0.156***	0.154***	0.149***	0.303**
	(0.007)	(0.009)	(0.010)	(0.154)	(0.006)	(0.008)	(0.008)	(0.133)
Observations	4,396	4,396	4,396	4,396	4,396	4,396	4,396	4,396
R-squared	0.001	0.001	0.002	0.049	0.001	0.001	0.002	0.033
Fixed Effects				YES				YES

Notes: The dependent variable for Column 1 through Column 4 is an indicator variable that equals 1 if the news article is about the shooter based on the keywords. Column 1 shows an OLS regression of an indicator on a dummy for hate-motivated mass shootings. Column 2 additionally controls for the number of dead victims and the number of injured victims in each shooting. Column 3 adds an interaction term between the number of dead victims and hate. Column 4 adds a set of fixed effects, including days since shooting fixed effect (from day 0, i.e., the day the shooting happened, to day 13, i.e., the 13th day since the shooting happened), day of the week fixed effect, month fixed effect, year fixed effect. Column 5 to Column 8 uses the same specifications as Columns 1 to Column 4, except that the dependent variable is identified based on the title of the news article. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.5: Percentage of News Articles about the Shooter's Motive

	Percentage of news articles about the shooter's motive			
	(1)	(2)	(3)	(4)
Hate	0.074***	0.071***	0.080***	0.079***
	(0.007)	(0.007)	(0.009)	(0.010)
#Dead		0.001	0.001**	0.001
		(0.000)	(0.001)	(0.001)
#Injured		-0.000	-0.000*	-0.000
		(0.000)	(0.000)	(0.000)
#Dead * Hate			-0.001	-0.001
			(0.001)	(0.001)
Constant	0.017***	0.014***	0.012***	0.114*
	(0.003)	(0.003)	(0.004)	(0.060)
Observations	4,396	4,396	4,396	4,396
R-squared	0.025	0.026	0.026	0.050
Fixed Effects				YES

Notes: The dependent variable is an indicator variable that equals 1 if the keywords of the news article contains "motive, ideology, manifesto, reason, racial, race, hate" Column 1 shows an OLS regression of an indicator on a dummy for hate-motivated mass shootings. Column 2 additionally controls for the number of dead victims and the number of injured victims in each shooting. Column 3 adds an interaction term between the number of dead victims and hate. Column 4 adds a set of fixed effects, including days since shooting fixed effect (from day 0, i.e., the day the shooting happened, to day 13, i.e., the 13th day since the shooting happened), day of the week fixed effect, month fixed effect, year fixed effect. Column 5 to Column 8 uses the same specifications as Columns 1 to Column 4, except that the dependent variable is identified based on the title of the news article. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Interest in Mass Shootings Measured by Online Searching Behavior

	State-level search interest on Google			
	(1)	(2)	(3)	(4)
Hate	12.733***	11.253***	7.130***	7.334***
	(0.581)	(0.576)	(0.755)	(0.790)
#Dead		0.474***	0.252***	0.404***
		(0.035)	(0.044)	(0.046)
#Injured		0.008	0.033***	0.019**
		(0.008)	(0.008)	(0.009)
#Dead * Hate			0.519***	0.340***
			(0.062)	(0.066)
Constant	7.629***	4.662***	5.799***	-2.260
	(0.212)	(0.272)	(0.302)	(1.894)
Observations	7,650	7,650	7,650	7,650
R-squared	0.059	0.103	0.111	0.170
Fixed Effects				YES

Notes: The dependent variable is the search interest value of a mass shooting in a subregion, as explained in Figure 1.4. Google divides the United States into 51 subregions based on geography. Thus, for each shooting, there are 51 observations. The data is restricted to reflect searching behaviors in the United States within 14 days since each shooting happened. Column 1 shows an OLS regression of search interest value on a dummy for hate-motivated mass shootings. Column 2 additionally controls for the number of dead victims and the number of injured victims in each shooting. Column 3 adds an interaction term between the number of dead victims and hate. Column 4 adds month fixed effect and year fixed effect. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.7: Interest in the Shooter by Online Searching Behavior

	=1 if "Suspect" is among the most searched topics			
	(1)	(2)	(3)	(4)
Hate	0.311***	0.294**	0.126	0.031
	(0.112)	(0.114)	(0.148)	(0.144)
#Dead		0.006	-0.004	0.006
		(0.007)	(0.009)	(0.008)
#Injured		-0.000	0.001	-0.001
		(0.002)	(0.002)	(0.002)
#Dead * Hate			0.021*	0.018
			(0.012)	(0.012)
Constant	0.289***	0.260***	0.307***	-0.347
	(0.042)	(0.055)	(0.060)	(0.275)
Observations	141	141	141	141
R-squared	0.053	0.058	0.079	0.401
Fixed Effects				YES

Notes: The dependent variable is an indicator variable that equals 1 if "Suspect" is in the list of the most searched topics related to the shooting. Column 1 shows an OLS regression of an indicator on a dummy for hate-motivated mass shootings. Column 2 additionally controls for the number of dead victims and the number of injured victims in each shooting. Column 3 adds an interaction term between the number of dead victims and hate. Column 4 adds a set of fixed effects, month fixed effect and year fixed effect. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.8: Interest in the Shooter's Motive by Online Searching Behavior

	=1 if motive is among the most searched topics			
	(1)	(2)	(3)	(4)
Hate	0.175***	0.170***	0.188***	0.181***
	(0.051)	(0.048)	(0.063)	(0.067)
#Dead		0.001	0.002	0.002
		(0.003)	(0.004)	(0.004)
#Injured		0.002***	0.002***	0.002***
		(0.001)	(0.001)	(0.001)
#Dead * Hate			-0.002	-0.004
			(0.005)	(0.006)
Constant	0.025	-0.001	-0.006	-0.100
	(0.019)	(0.023)	(0.026)	(0.128)
Observations	141	141	141	141
R-squared	0.079	0.216	0.217	0.388
Fixed Effects				YES

Notes: The dependent variable is an indicator variable that equals 1 if the most searched topics contains any of the following: "motive, ideology, manifesto, reason, racial, race, hate". Column 1 shows an OLS regression on a dummy for hate-motivated mass shootings. Column 2 additionally controls for the number of dead victims and the number of injured victims in each shooting. Column 3 adds an interaction term between the number of dead victims and hate. Column 4 adds month fixed effect and year fixed effect. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Recruitment

Treatment	**Platform**	**# Subjects**
T1: No Hate	Prolific	402
	CloudResearch	200
T2: Hate	Prolific	399
	CloudResearch	202
T3: Hate Ideology	Prolific	399
	CloudResearch	198
T4: Hate Background	Prolific	401
	CloudResearch	199

Table A.10: Summary Statistics and Balance Tests

	A: By political affiliation			B: By recruiting platform		
	Democrat (1)	Republican (2)	t-test (1)=(2)	Prolific (3)	CloudResearch (4)	t-test (3)=(4)
Age	37.81	38.70	0.10*	37.48	39.82	0.00***
	(12.95)	(13.35)		(13.07)	(13.20)	
Education level	3.82	3.70	0.01***	3.79	3.70	0.09*
	(1.08)	(1.12)		(1.14)	(1.04)	
Income level	6.75	7.48	0.00***	7.36	6.62	0.00***
	(3.33)	(3.35)		(3.39)	(3.25)	
White	0.69	0.79	0.00***	0.75	0.72	0.11
	(0.46)	(0.41)		(0.43)	(0.45)	
Political index	0.08	0.64	0.00***	0.35	0.38	0.05**
	(0.14)	(0.26)		(0.35)	(0.36)	
Fame-seeking index	2.56	2.85	0.00***	2.69	2.74	0.27
	(0.89)	(0.84)		(0.86)	(0.90)	
Observations	1,199	1,201		1,601	799	

Notes: This table reports the mean of each demographic variable for the Democrat sample (1), the Republican sample (2), the Prolific sample (3), and the CloudResearch sample (4). The corresponding standard deviation is reported in parentheses. t-test of inequality between (1) and (2), and t-test of inequality between (3) and (4) are reported in the last column of each panel. Education level is a categorical variable ranging from Less than high school (1) to Doctorate (7). Income level is a categorical variable ranging from Less than $10,000 (1) to $15,000 or more (12). Political index ranges from 0 to 1, a higher value means more right-leaning. Fame-seeking index ranges from 1 to 5, a higher value means more fame-seeking.

Table A.11: Treatment Effects on Support for the Shooter

	A: Full Sample			B: Democrat			C: Republican		
	Admire (1)	Justify (2)	Sentence (3)	Admire (4)	Justify (5)	Sentence (6)	Admire (7)	Justify (8)	Sentence (9)
No Hate (T1)	0.123***	0.113**	0.212***	0.077*	0.056	0.201***	0.175***	0.178**	0.246***
	(0.039)	(0.046)	(0.059)	(0.044)	(0.055)	(0.073)	(0.064)	(0.072)	(0.090)
Hate Ideology (T3)	0.086**	0.054	0.136**	0.013	-0.001	0.059	0.155**	0.112	0.226**
	(0.039)	(0.046)	(0.059)	(0.044)	(0.055)	(0.073)	(0.064)	(0.072)	(0.090)
Hate Background (T4)	0.058	0.061	0.177***	0.124***	0.139**	0.252***	0.003	0.005	0.121
	(0.039)	(0.046)	(0.059)	(0.044)	(0.055)	(0.073)	(0.064)	(0.073)	(0.091)
Control Mean	1.126	1.250	1.781	1.067	1.150	1.906	1.186	1.349	1.657
Observations	2,400	2,400	2,383	1,199	1,199	1,190	1,201	1,201	1,193
R-squared	0.049	0.048	0.084	0.147	0.091	0.063	0.044	0.065	0.118
T1=T3 p-value	0.347	0.207	0.197	0.152	0.306	0.052*	0.751	0.362	0.827
T1=T4 p-value	0.097*	0.268	0.549	0.287	0.133	0.476	0.007***	0.017**	0.169
T3=T4 p-value	0.473	0.878	0.492	0.013**	0.012**	0.008***	0.017**	0.140	0.246
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sub-sample. Panel C shows the results for the Republican sub-sample. The dependent variables in each panel are respectively: admiration for the shooter (5 point Likert-scale), justification for the shooter's action (5 point Likert-scale), and sentencing option for the shooter (6 options ranging from 10 years or less imprisonment to death penalty). The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.12: Heterogeneous Treatment Effect by Political Affiliations

	Demand manifesto		Demand background		Index support		Donation anti-immigrant		Hate links requested		Hate links clicked	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
No Hate (T1)	0.055	0.046	0.058	0.055	0.190**	0.140*	0.028	0.020	0.016	0.007	-0.003	-0.004
	(0.040)	(0.040)	(0.040)	(0.040)	(0.083)	(0.081)	(0.039)	(0.038)	(0.025)	(0.025)	(0.008)	(0.008)
Hate Ideology (T3)	-0.016	-0.026	-0.001	-0.004	0.058	0.028	0.020	0.014	0.018	0.011	0.003	0.003
	(0.040)	(0.040)	(0.040)	(0.040)	(0.084)	(0.081)	(0.038)	(0.038)	(0.025)	(0.025)	(0.008)	(0.008)
Hate Background (T4)	-0.031	-0.033	-0.135***	-0.133***	0.282***	0.243***	0.071*	0.061	0.040	0.034	0.003	0.003
	(0.040)	(0.040)	(0.040)	(0.040)	(0.083)	(0.081)	(0.038)	(0.037)	(0.025)	(0.025)	(0.008)	(0.008)
Republican (Rep)	0.058	0.003	-0.012	-0.041	0.030	0.008	0.163***	0.043	0.063**	0.022	0.010	0.001
	(0.040)	(0.048)	(0.040)	(0.049)	(0.083)	(0.098)	(0.038)	(0.046)	(0.025)	(0.030)	(0.008)	(0.010)
No Hate * Rep	0.005	0.009	0.002	0.002	0.136	0.178	-0.060	-0.053	-0.003	0.004	-0.007	-0.006
	(0.057)	(0.056)	(0.057)	(0.057)	(0.118)	(0.114)	(0.055)	(0.054)	(0.036)	(0.035)	(0.012)	(0.012)
Hate Ideology * Rep	-0.132**	-0.119**	-0.110*	-0.107*	0.171	0.217*	-0.065	-0.060	-0.020	-0.011	0.007	0.007
	(0.057)	(0.056)	(0.057)	(0.057)	(0.118)	(0.114)	(0.054)	(0.053)	(0.036)	(0.035)	(0.012)	(0.012)
Hate Background * Rep	-0.006	-0.006	0.018	0.017	-0.269**	-0.181	-0.172***	-0.155***	-0.056	-0.043	-0.017	-0.016
	(0.057)	(0.056)	(0.057)	(0.057)	(0.118)	(0.115)	(0.053)	(0.053)	(0.036)	(0.035)	(0.012)	(0.012)
Control Mean (Hate & Dem)	0.407	0.407	0.563	0.563	0.001	0.001	0.114	0.114	0.067	0.067	0.007	0.007
Observations	2,400	2,400	2,400	2,400	2,400	2,400	1,665	1,665	2,400	2,400	2,400	2,400
R-squared	0.015	0.034	0.023	0.028	0.015	0.077	0.019	0.050	0.006	0.040	0.005	0.010
T1=T3 p-value	0.076	0.069	0.145	0.142	0.112	0.166	0.835	0.869	0.965	0.870	0.414	0.407
T1=T4 p-value	0.031	0.047	0.000	0.000	0.272	0.204	0.250	0.267	0.355	0.286	0.422	0.409
T3=T4 p-value	0.706	0.871	0.001	0.002	0.007	0.008	0.167	0.196	0.380	0.369	0.987	0.995
T1+T1*Rep=0 p-value	0.135	0.167	0.138	0.155	0.000	0.000	0.399	0.396	0.597	0.659	0.228	0.227
T3+T3*Rep=0 p-value	0.000	0.000	0.006	0.006	0.006	0.002	0.230	0.226	0.908	0.982	0.224	0.214
T4+T4*Rep=0 p-value	0.356	0.331	0.004	0.004	0.884	0.447	0.008	0.013	0.529	0.704	0.109	0.119
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table presents results from OLS regressions. The dependent variables from left to right are: (1) an indicator variable that equals 1 if the subject requested to be shown the shooter's manifesto, (2) an indicator variable that equals 1 if the subject requested to be shown the shooter's Background, (3) a standardized index measuring support for the shooter, (4) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (5) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (6) a dummy that equals 1 if the subject clicked on the provided links about the hate group. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment, a dummy for Republican subjects, and the interaction terms between treatment assignment and party affiliation. Democrat subjects in the Hate Treatment is the omitted group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.13: Heterogeneous Treatment Effects within Sub-samples

	A: Democrat sample				B: Republican sample			
	Index Support (1)	Donation anti-immigrant (2)	Links requested (3)	Links clicked (4)	Index Support (5)	Donation anti-immigrant (6)	Links requested (7)	Links clicked (8)
No Hate (T1)	0.115	0.024	0.016	-0.008	0.605***	0.059	0.054	-0.032
	(0.076)	(0.038)	(0.025)	(0.008)	(0.193)	(0.122)	(0.078)	(0.027)
Hate Ideology (T3)	0.010	0.011	0.018	0.005	0.091	-0.036	-0.053	-0.009
	(0.076)	(0.037)	(0.025)	(0.008)	(0.191)	(0.125)	(0.077)	(0.027)
Hate Background (T4)	0.138*	0.057	0.004	-0.004	0.003	-0.152	-0.048	-0.034
	(0.076)	(0.038)	(0.025)	(0.008)	(0.203)	(0.131)	(0.081)	(0.028)
Unfriendly	0.051	0.110*	0.027	-0.012	-0.046	-0.010	0.016	-0.016
	(0.122)	(0.063)	(0.040)	(0.013)	(0.154)	(0.099)	(0.062)	(0.021)
No Hate * Unfriendly	0.198	-0.043	-0.030	0.024	-0.426**	-0.116	-0.049	0.023
	(0.169)	(0.085)	(0.055)	(0.018)	(0.206)	(0.130)	(0.083)	(0.029)
Hate Ideology * Unfriendly	0.142	0.028	-0.020	-0.004	0.113	-0.011	0.065	0.022
	(0.169)	(0.086)	(0.055)	(0.018)	(0.204)	(0.132)	(0.082)	(0.028)
Hate Background * Unfriendly	0.521***	0.027	0.151***	0.035**	0.046	0.065	0.046	0.023
	(0.168)	(0.083)	(0.055)	(0.017)	(0.215)	(0.138)	(0.086)	(0.030)
Control Mean	0.000	0.114	0.067	0.007	0.001	0.278	0.130	0.017
Observations	1,199	842	1,199	1,199	1,201	823	1,201	1,201
R-squared	0.083	0.069	0.062	0.013	0.118	0.034	0.039	0.013
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the Democrat sample. Panel B shows the results for the Republican sample. The dependent variables in each panel from left to right are: (1) a standardized index measuring support for the shooter, (2) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (3) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (4) a dummy that equals 1 if the subject clicked on the provided links about the hate group. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, a dummy for the Hate Background Treatment, Political Index (PI), and interaction terms between the treatment dummies and the Political Index. The Hate Treatment is the omitted group. Unfriendly is an indicator that equals to 1 if the subject support building a wall along the U.S. southern border, or does not support children of illegal immigrants to be granted legal citizenship. Control variables include age, income, education, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and in indicator variable that equals 1 if the subject is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.14: Experimenter Demand Effect

	All	T1	T2	T3	T4	T1=T2	T3=T2	T4=T2
Immigrant	0.08	0.04	0.08	0.12	0.07	0.01***	0.013**	0.748
Race	0.06	0.02	0.05	0.08	0.07	0.006***	0.045***	0.268
Hate	0.06	0.02	0.09	0.07	0.07	0.00***	0.308	0.167
One of the above three	0.18	0.08	0.21	0.25	0.20	0.000***	0.114	0.732

Notes: This table reports the percentage of subjects who are able to correctly guess the purpose of the experiment. The first row reports the percentage of subjects whose response is related to immigrant. The second row reports the percentage of subjects whose response is related to race. The third row reports the percentage of subjects whose response is related to hate. The last row reports the percentage of subjects whose response is related to immigrant or race or hate. The last three columns report p-value from t-test.

Table A.15: Robustness Checks

	A: Full Sample			B: Democrat			C: Republican		
	Guessed purpose correctly	Passed all attention checks	Recognize the shooting	Guessed purpose correctly	Passed all attention checks	Recognize the shooting	Guessed purpose correctly	Passed all attention checks	Recognize the shooting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
No Hate (T1)	-0.123***	0.019	-0.053**	-0.143***	0.023	-0.069*	-0.106***	0.019	-0.034
	(0.022)	(0.020)	(0.025)	(0.031)	(0.023)	(0.036)	(0.032)	(0.031)	(0.034)
Hate Ideology (T3)	0.040*	0.012	-0.002	-0.018	0.007	-0.019	0.099***	0.020	0.019
	(0.022)	(0.020)	(0.025)	(0.031)	(0.023)	(0.036)	(0.031)	(0.031)	(0.034)
Hate Background (T4)	-0.008	0.051***	0.036	-0.056*	0.016	0.047	0.030	0.088***	0.027
	(0.022)	(0.020)	(0.025)	(0.031)	(0.023)	(0.036)	(0.032)	(0.031)	(0.035)
Control Mean	0.208	0.844	0.256	0.230	0.897	0.280	0.186	0.791	0.233
Observations	2,400	2,400	2,400	1,199	1,199	1,199	1,201	1,201	1,201
R-squared	0.038	0.034	0.021	0.035	0.024	0.037	0.053	0.041	0.012
T1=T3 p-value	0.000	0.737	0.040	0.000	0.499	0.161	0.000	0.971	0.130
T1=T4 p-value	0.000	0.100	0.000	0.004	0.787	0.001	0.000	0.028	0.081
T3=T4 p-value	0.027	0.048	0.120	0.221	0.685	0.063	0.031	0.031	0.814
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents results from OLS regressions. Panel A shows the results for the full sample. Panel B shows the results for the Democrat sample. Panel C shows the results for the Republican sample. The dependent variables in each panel from left to right are: (1) a dummy that equals 1 if the subject correctly guessed the purpose of the experiment, (2) a dummy that equals 1 if the subject passed both attention checks, (3) A dummy that equals 1 if the subject recognized the shooting. The independent variables include a dummy for the No Hate Treatment, a dummy for the Hate Ideology Treatment, and a dummy for the Hate Background Treatment. The Hate Treatment is the omitted group. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.16: The Correlation between Social Desirability Bias and Outcome Measures

	Index support		Donation anti-immigrant		Hate links requested		Hate links clicked	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SD Index	-0.207	-0.249	-0.008	-0.021	-0.072	-0.100	-0.017	-0.021
	(0.181)	(0.182)	(0.096)	(0.095)	(0.063)	(0.063)	(0.014)	(0.014)
Constant	0.292**	0.091	0.198***	-0.157	0.163***	0.025	0.016	-0.037*
	(0.128)	(0.269)	(0.069)	(0.138)	(0.044)	(0.094)	(0.010)	(0.021)
Observations	399	399	280	280	399	399	399	399
R-squared	0.003	0.058	0.000	0.059	0.003	0.043	0.004	0.034
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table presents results from OLS regressions. SD Index is an index ranging from 0 to 1 that measures a subject's propensity to give the socially desirable response. The dependent variables in each panel from left to right are: (1) a standardized index measuring support for the shooter, (2) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (3) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (4) a dummy that equals 1 if the subject clicked on the provided links about the hate group. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.17: Correlation between Attitudinal Measures and Behavioral Outcome

	A: Full sample			B: Democrat			C: Republican		
	Donation anti-immigrant	Hate links Requested	Hate links clicked	Donation anti-immigrant	Hate links Requested	Hate links clicked	Donation anti-immigrant	Hate links Requested	Hate links clicked
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Index of support	0.105***	0.083***	0.013***	0.110***	0.071***	0.016***	0.095***	0.084***	0.010***
	(0.011)	(0.007)	(0.002)	(0.016)	(0.011)	(0.004)	(0.016)	(0.010)	(0.004)
Constant	-0.021	0.031	-0.009	0.013	0.017	-0.015	-0.053	0.069	-0.005
	(0.048)	(0.032)	(0.011)	(0.057)	(0.039)	(0.013)	(0.089)	(0.057)	(0.020)
Observations	1,665	2,400	2,400	842	1,199	1,199	823	1,201	1,201
R-squared	0.092	0.089	0.017	0.129	0.086	0.028	0.066	0.090	0.013
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

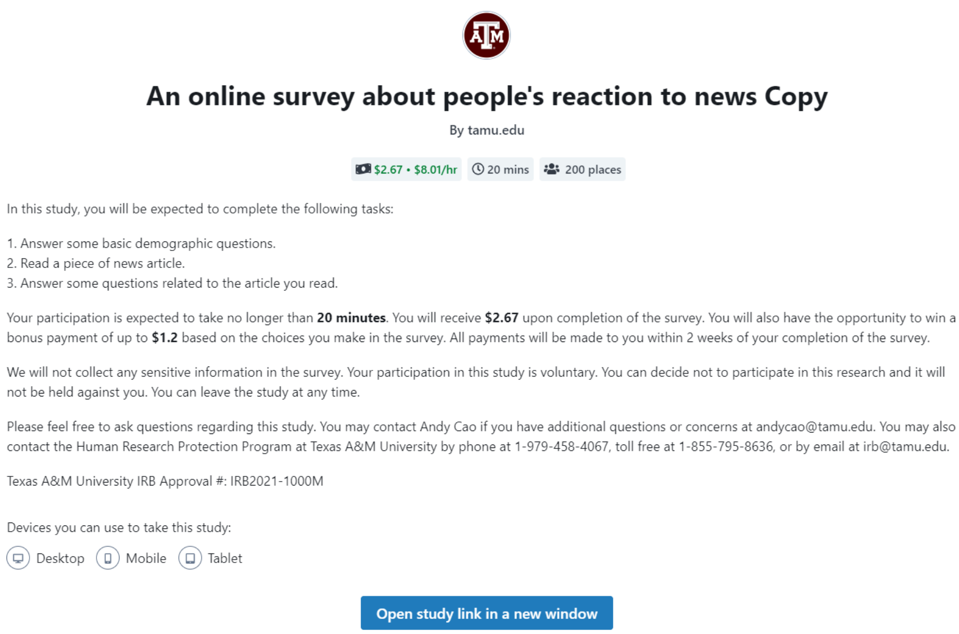
Notes: This table presents results from OLS regressions. The dependent variables in each panel from left to right are: (1) a dummy that equals 1 if the subject authorized the $1 donation to the anti-immigrant organization, (2) a dummy that equals 1 if the subject requested to be shown links to access the website of a white supremacy hate group, (3) a dummy that equals 1 if the subject clicked on the provided links about the hate group. The independent variable is the standardized index of support for the shooter. Control variables include age, income, education, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and in indicator variable that equals 1 if the subject is recruited from CloudResearch. Estimates with controls are reported in odd columns. Estimates without controls are reported in even columns. Control variables include age, income, education, an index measuring political stance, an index measuring fame-seeking personality, an indicator variable that equals 1 if the subject is white, and an indicator variable that equals 1 if the subjects is recruited from CloudResearch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A.2 Appendix B Survey Materials

Appendix A.2.1 shows the recruiting advertisement of the study. Appendix A.2.2 shows each of the four information treatments. Appendix A.2.3 shows my main outcome measures. Appendix A.2.4 shows other survey questions mentioned in the paper. To view the complete survey interactively, please visit `https://tamu.qualtrics.com/jfe/form/SV_e4pM5zr0oXP3Xvg`

A.2.1 Advertisement

Figure A.2: Study Description on Prolific



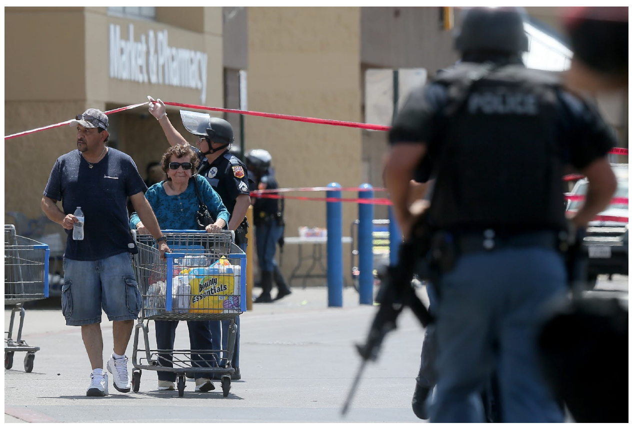
A.2.2 Information Treatments

A.2.2.1 Treatment 1: No Hate

A gunman opened fire Saturday in a grocery store and around a nearby shopping mall, leaving 20 people dead and 26 injured. The suspect is in custody and police say no officers fired their weapons while arresting him.

The police and FBI are investigating whether an anonymous “manifesto,” shared on an online forum, was written by the gunman. The document claims that the attack was motivated by economic reasons.

The store where the shooting took place is a popular destination among tourists. The Department of Justice has called the shooting an act of domestic terrorism. One wit-



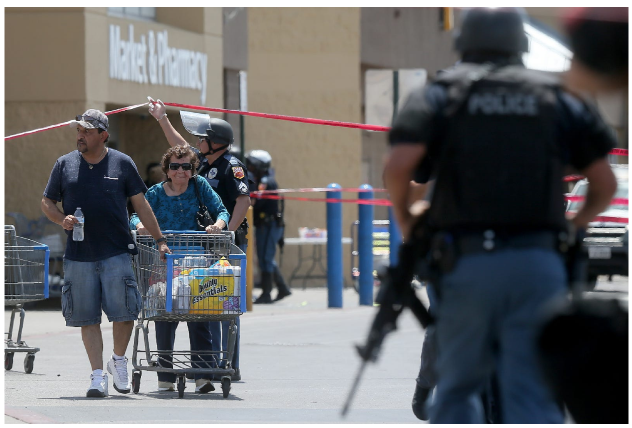
ness said she was shopping with her husband when they heard gunfire. “People were panicking and running, saying that there was a shooter,” “They were running close to the floor, people were dropping on the floor.” She and her husband ran through a stock room before taking cover with other customers.

A.2.2.2 Treatment 2: Hate

A gunman opened fire Saturday in a grocery store and around a nearby shopping mall, leaving 20 people dead and 26 injured. The suspect is in custody and police say no officers fired their weapons while arresting him.

The victims include at least 8 Mexicans, law enforcement officials said. The police and FBI are investigating whether an anonymous white nationalist “manifesto,” shared on an online forum, was written by the gunman. The document claims that the attack was targeted at the local Hispanic community.

The store where the shooting took place is a popular destination among Mexican tourists. The Department of Justice has called the shooting an act of domestic terrorism and federal authorities say they are investigating possible hate crimes charges. One witness said she was shopping with her husband when they heard gunfire. “Peo-



ple were panicking and running, saying that there was a shooter,” “They were running close to the floor, people were dropping on the floor.” She and her husband ran through a stock room before taking cover with other customers.

A.2.2.3 Treatment 3: Hate Ideology

A gunman opened fire Saturday in a grocery store and around a nearby shopping mall, leaving 20 people dead and 26 injured. The suspect is in custody and police say no officers fired their weapons while arresting him.

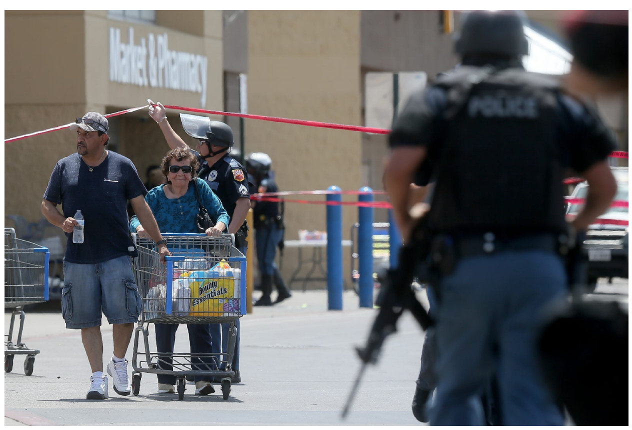
The victims include at least 8 Mexicans, law enforcement officials said. The police and FBI are investigating whether an anonymous white nationalist “manifesto,” shared on an online forum, was written by the gunman.

The document claims that the attack was targeted at the local Hispanic community. It stated that Latin America immigrants represented a "Hispanic invasion." It warned that white people were being replaced by foreigners.

The manifesto described an imminent attack and railed against immigrants, saying, "if we can get rid of enough people, then our way of life can be more sustainable." It also detailed a plan to separate America into territories by race to save this country.

The author hoped his/her attack and words would inspire additional like-minded attacks and lead to a wider racial violence in pursuit of a white ethnostate.

The store where the shooting took place is a popular destination among Mexican tourists. The Department of Justice has called the shooting an act of domestic terrorism and federal authorities say they are investigating possible hate crimes charges. One witness said she was shopping with her husband when they heard gunfire. "Peo-



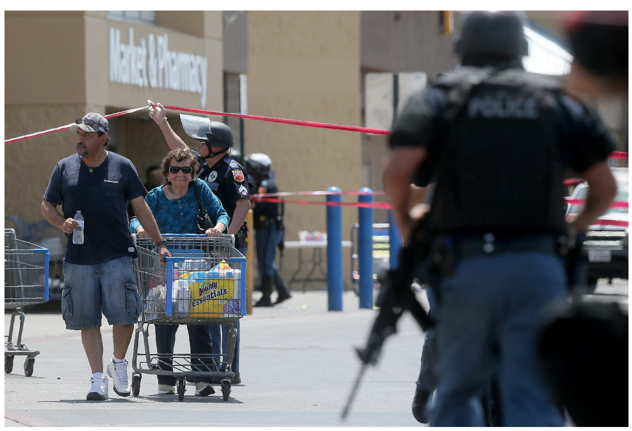
ple were panicking and running, saying that there was a shooter," "They were running close to the floor, people were dropping on the floor." She and her husband ran through a stock room before taking cover with other customers.

A.2.2.4 Treatment 4: Hate Background

A gunman opened fire Saturday in a grocery store and around a nearby shopping mall, leaving 20 people dead and 26 injured. The suspect is in custody and police say no officers fired their weapons while arresting him.

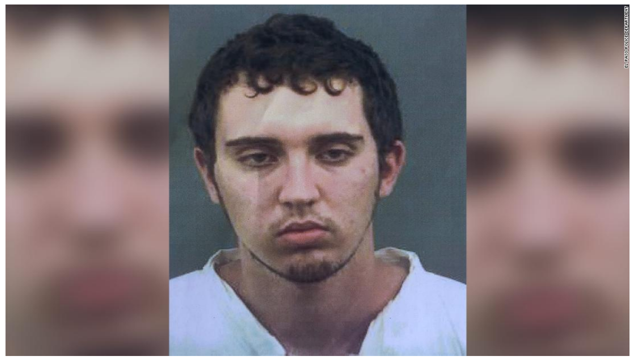
The victims include at least 8 Mexicans, law enforcement officials said. The police and FBI are investigating whether an anonymous white nationalist “manifesto,” shared on an online forum, was written by the gunman. The document claims that the attack was targeted at the local Hispanic community.

The store where the shooting took place is a popular destination among Mexican tourists. The Department of Justice has called the shooting an act of domestic terrorism and federal authorities say they are investigating possible hate crimes charges. One witness said she was shopping with her husband when they heard gunfire. “Peo-



ple were panicking and running, saying that there was a shooter,” “They were running close to the floor, people were dropping on the floor.” She and her husband ran through a stock room before taking cover with other customers.

Police officers were interviewing the suspect, Patrick Crusius, a 21-year-old white man from Allen, Tex. Investigators are looking into whether Crusius might have been



radicalized online. But friends and former teachers and classmates say he might have been hardened, too, by the tensions in his changing community in real life.

Allison Pettitt, a classmate, said she saw Crusius pushed around in the hallways and "cussed out by some of the Spanish-speaking kids." She said that bullying was common at the school and that teachers often ignored it. "He started getting more depressed closer to the end of junior year," Pettitt said. "He started wearing a trench coat to school and becoming more antisocial and withdrawn." Lesley Range-Stanton, a spokeswoman for Plano's school district, declined to comment about whether Crusius was bullied, citing student privacy.

A.2.3 Outcome measures

A.2.3.1 Interest in the Shooter

The news story mentioned an anonymous manifesto. The manifesto was published on an online forum 19 minutes before the first 911 call that alerted the authorities to the mass shooting at the mall.

Would you like to read the anonymous manifesto? If you choose “Yes” we will provide you with access to the complete manifesto at the end of the survey. If you choose “No” you will proceed with the survey without receiving access.

- Yes
- No

Since the shooting happened, authorities have released more information about the suspect of the shooting.

Would you like to know more about the suspect? E.g., age, background. If you choose “Yes” we will provide you with more information at the end of the survey. If you choose “No” you will proceed with the survey without receiving information.

- Yes
- No

A.2.3.2 Justification for the Shooter (No Hate)

Justification is a defense in a criminal case, by which a defendant who committed the crime as defined, claims they did no wrong, because committing the crime advanced some social interest or vindicated a right of such importance that it outweighs the wrongfulness of the crime.
Now think about the news story you just read. The police has confirmed that the suspect was the author of the manifesto, which claims that the attack was motivated by economic reasons.
On a scale of 5, do you think the suspect’s reason for the attack satisfies the justification defense?

A.2.3.3 Justification for the Shooter (Hate)

Justification is a defense in a criminal case, by which a defendant who committed the crime as defined, claims they did no wrong, because committing the crime advanced some social interest or vindicated a right of such importance that it outweighs the wrongfulness of the crime.

Now think about the news story you just read. The police has confirmed that the suspect was the author of the manifesto, which claims that the attack was targeted at the local Hispanic community.

On a scale of 5, do you think the suspect's reason for the attack satisfies the justification defense?

A.2.3.4 Sentencing for the Shooter

Now think about the other people who are taking this survey. What do you think is the sentencing option that most other people would choose?

- 10 years or less imprisonment
- 11 to 20 years imprisonment
- 21 to 39 years imprisonment
- Life imprisonment with parole after a minimum of 30 years
- Life imprisonment without parole
- Death Penalty
- Other, please specify

A.2.3.5 Admiration for the Shooter

In your own opinion, how much do you agree or disagree with the following statement?
"Even though what the shooter did was wrong, it took a lot of courage to do what he did, and I admire him for that."

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

A.2.3.6 Support for the Shooter's Ideology (anti-immigrant)

As part of the study, we will now randomly select an organization. You will have the opportunity to authorize a $1 donation to that organization. You will learn about the organization in the next page.

Your decision will NOT affect your payoff from the survey. If you choose to authorize the $1 donation, this money will NOT be deducted from your payment, we will simply make the donation on your behalf.

We will post the total amount of donation made on our website once the study is published. All results reported will be in an aggregate format, your individual decision is strictly anonymous. If you are interested, you can visit this website for updates.

> The organization randomly selected for you is the Federation for American Immigration Reform (FAIR).
>
> The Federation for American Immigration Reform (FAIR) is a non profit, anti-immigration organization in the United States. The group publishes position papers, organizes events, and runs campaigns in order to advocate for changes in U.S. immigration policy.
>
> FAIR leaders have ties to white supremacist groups and eugenicists and have made many anti-immigration statements. FAIR's founder, John Tanton, has expressed his wish that America remain a majority-white population: a goal to be achieved, presumably, by limiting the number of nonwhites who enter the country.

Would you like to authorize a $1 donation to FAIR?

A.2.3.7 Support for the Shooter's Ideology (pro-immigrant)

As part of the study, we will now randomly select an organization. You will have the opportunity to authorize a $1 donation to that organization. You will learn about the organization in the next page.

Your decision will NOT affect your payoff from the survey. If you choose to authorize the $1 donation, this money will NOT be deducted from your payment, we will simply make the donation on your behalf.

We will post the total amount of donation made on our website once the study is published. All results reported will be in an aggregate format, your individual decision is strictly anonymous. If you are interested, you can visit this website for updates.

> The organization randomly selected for you is theAmerican Immigration Council (AIC).
>
> The American Immigration Council is a Washington, D.C.-based 501(c)(3) nonprofit organization and advocacy group. It advocates for immigrants to the United States.
>
> The American Immigration Council works to strengthen America by shaping how America thinks about and acts towards immigrants and immigration and by working toward a more fair and just immigration system that opens its doors to those in need of protection and unleashes the energy and skills that immigrants bring.

Would you like to authorize a $1 donation to AIC?

A.2.3.8 Interest in a White-supremacy Hate Group

There are organizations/groups whose goals and activities are primarily or substantially based on a shared antipathy towards people of one or more other different races, religions, ethnicities/nationalities/national origins, genders, and/or sexual identities.

Some of these organizations/groups share similar ideology as the author of the manifesto.

One example of such group is Stormfront. Created by former Alabama Klan boss and long-time white supremacist Don Black in 1995, Stormfront was the first major hate site on the Internet. Claiming more than 300,000 registered members as of May 2015 (though far fewer remain active), the site has been a very popular online forum for white nationalists. In its own words, “Our mission is to provide information not available in the controlled news media and to build a community of White activists working for the survival of our people.”

Stormfront was filtered out from Google Search due to its controversial content. Would you like to know how to access its website?

If you choose “Yes”, we will provide you with relevant information at the end of the survey. If you choose “No”, you will proceed with the survey without receiving access.

A.2.3.9 Example of Norm Elicitation

Earlier this month, 200 participants on MTurk participated in this survey. We asked these participants the same questions as the ones you are answering, including the last question you just saw: “On a scale of 1 to 5, do you think the suspect’s reason for the attack satisfies the justification defense?”

Now, we ask you to think about the answers of these 200 previous participants and guess what option most of them chose. You will win a bonus payment of $0.2 if your guess is correct.

200 participants were asked: “On a scale of 1 to 5, do you think the suspect’s reason for the attack satisfies the justification defense?”

I think most people chose:

A.2.4 Other Survey Questions

A.2.4.1 Measuring Political Stance

- What is your stance on abortion?
  - Pro-life
  - Pro-choice
- Do you support the legalization of same sex marriage?
  - Yes
  - No
- Should there be more restrictions on the current process of purchasing a gun?
  - Yes

- No

- What is your stance on abortion?
  - Pro-life
  - Pro-choice

- Should the government raise the federal minimum wage?
  - Yes
  - No

- Should the U.S. build a wall along the southern border?
  - Yes
  - No

- Should children of illegal immigrants be granted legal citizenship?
  - Yes
  - No

A.2.4.2 Measuring Social Desirability Bias

- Do you agree or disagree with the following statement? I'm always willing to admit it when I make a mistake
  - Agree
  - Disagree

- Do you agree or disagree with the following statement? I like to gossip at times
  - Agree
  - Disagree

- Do you agree or disagree with the following statement? There have been occasions when I took advantage of someone

- Agree
- Disagree

• Do you agree or disagree with the following statement? I sometimes try to get even rather than forgive and forget

  - Agree
  - Disagree

• Do you agree or disagree with the following statement? At times I have really insisted on having things my own way

  - Agree
  - Disagree

• Do you agree or disagree with the following statement? I have never been irked when people expressed ideas very different from my own

  - Agree
  - Disagree

• Do you agree or disagree with the following statement? I have never deliberately said something that hurt someone's feelings

  - Agree
  - Disagree

A.2.4.3 Example of Attention Checks

Thank you again for participating in our study. It is very important that you read and answer each question carefully. To show that you are paying attention, please choose both "Extremely displeased" and "slightly pleased" on the question below.

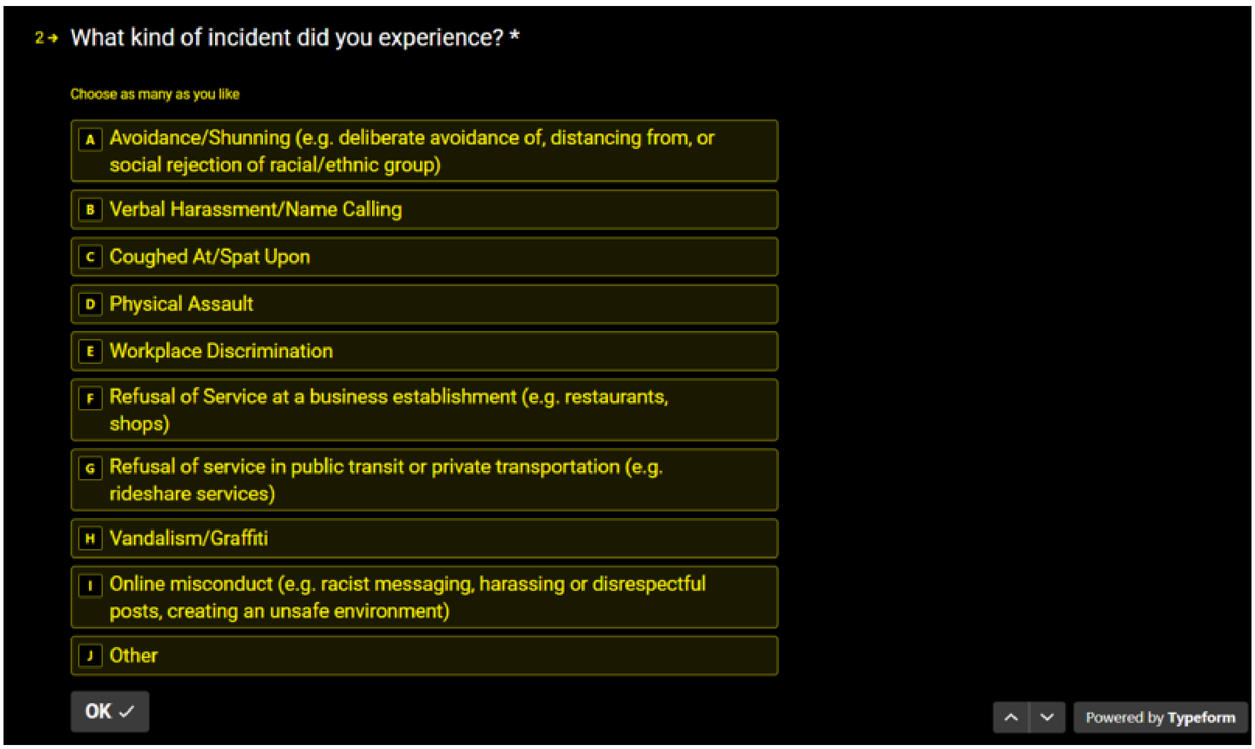
• How pleased are you with the weather today?

- Extremely displeased
- Moderately displeased
- Slightly displeased
- Neither pleased nor displeased
- Slightly pleased
- Moderately pleased
- Extremely pleased

APPENDIX B

CHAPTER II APPENDIX

Figure B.1: Stop AAPI HATE webpage (stopaaiphate.org), Last Accessed 10/4/2021



2→ What kind of incident did you experience? *

Choose as many as you like

A Avoidance/Shunning (e.g. deliberate avoidance of, distancing from, or social rejection of racial/ethnic group)

B Verbal Harassment/Name Calling

C Coughed At/Spat Upon

D Physical Assault

E Workplace Discrimination

F Refusal of Service at a business establishment (e.g. restaurants, shops)

G Refusal of service in public transit or private transportation (e.g. rideshare services)

H Vandalism/Graffiti

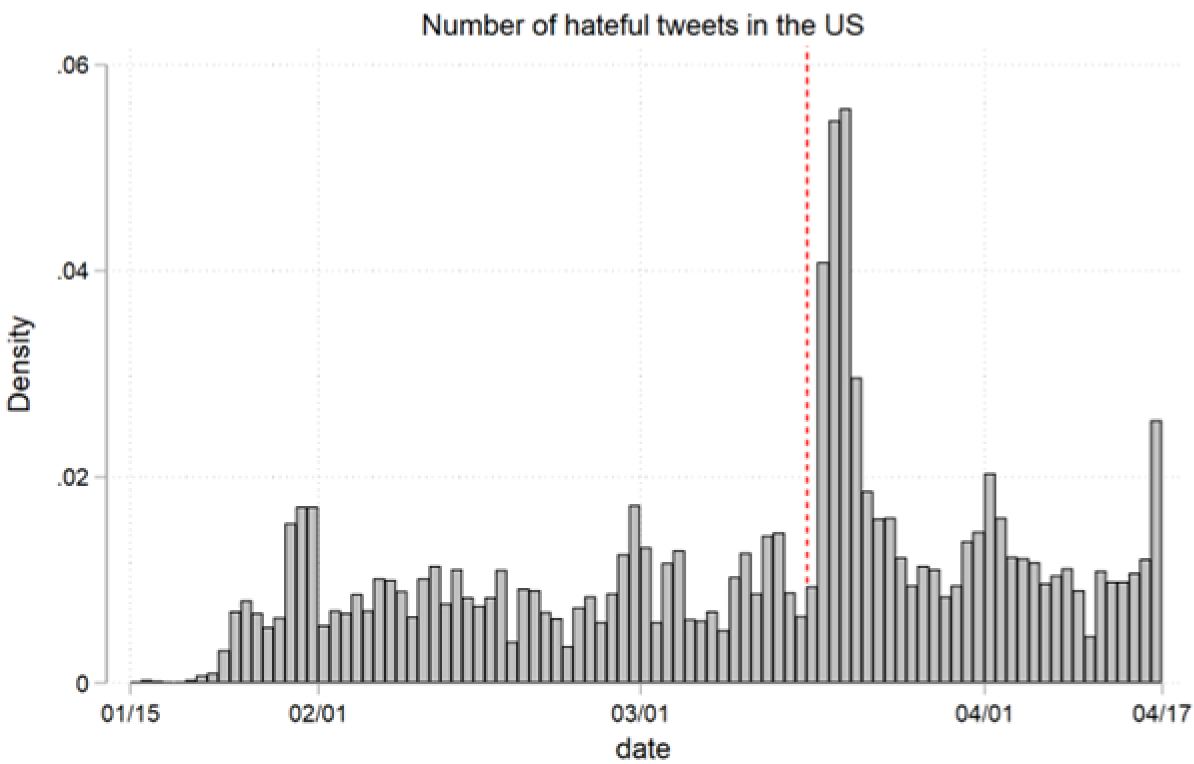
I Online misconduct (e.g. racist messaging, harassing or disrespectful posts, creating an unsafe environment)

J Other

OK ✓

Powered by Typeform

Figure B.2: Daily Anti-Asian COVID-19 Posts on Twitter Relative to Trump's First Public Use of "Chinese Virus"



Notes: Data are from He et al. (2021), which identifies 40,606 anti-Asian hate tweets from U.S. users.

Figure B.3: Daily Anti-Asian Incidents, Adjusted by Day-of-week Averages

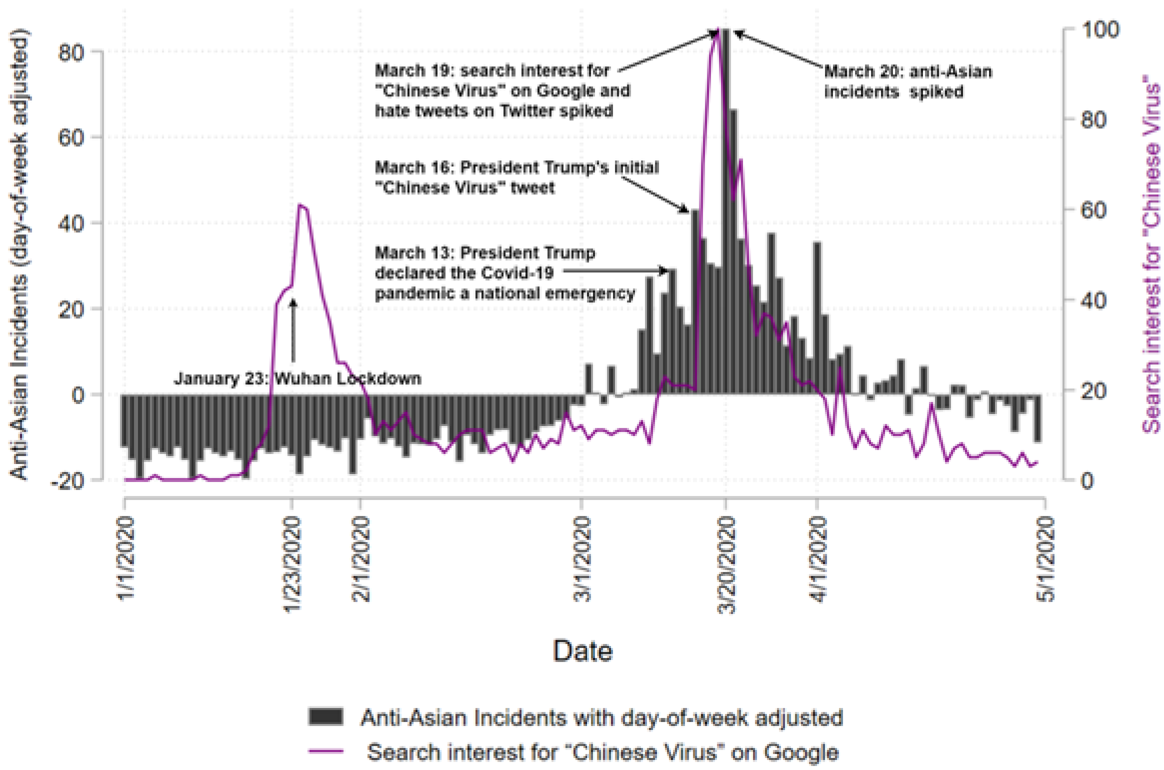
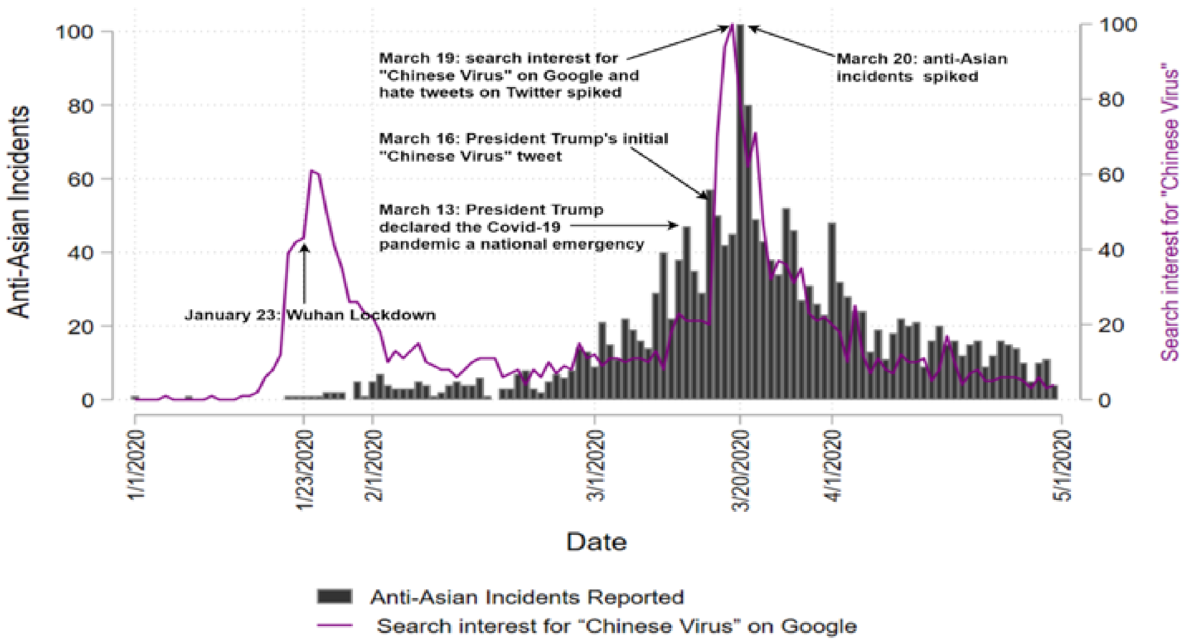


Figure B.4: Daily Anti-Asian Incidents by County Urbanicity

Panel A: Urban areas (county population > 150,000)



Panel B: Rural areas (county population < 150,000)

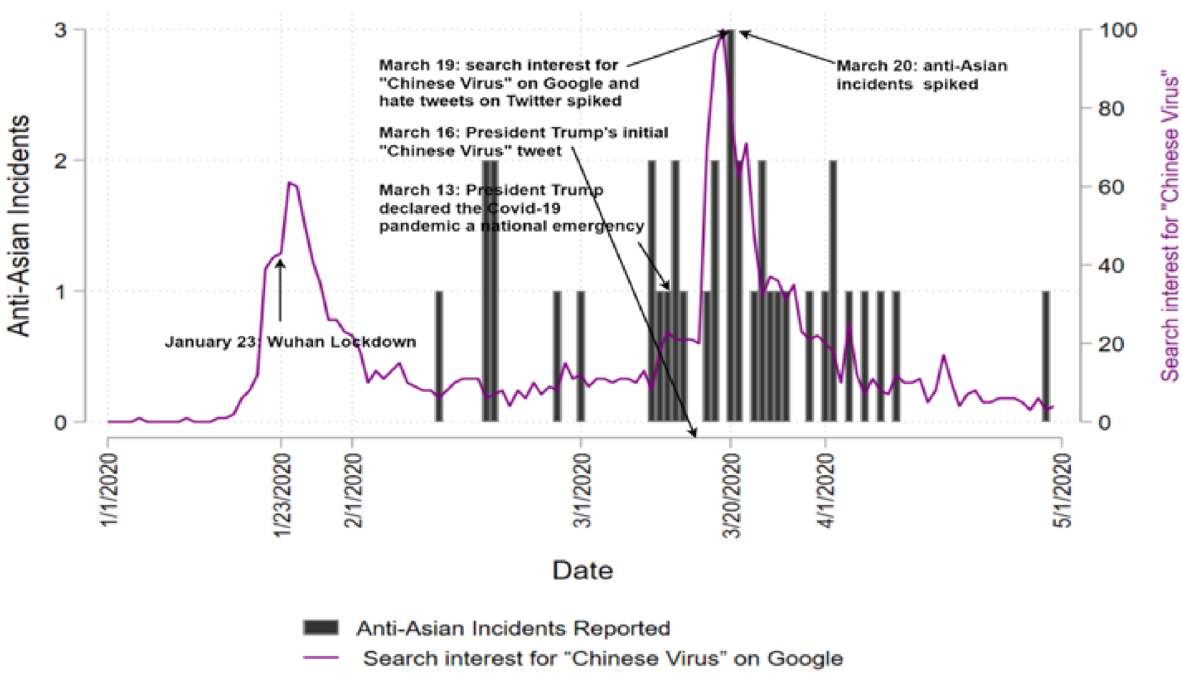
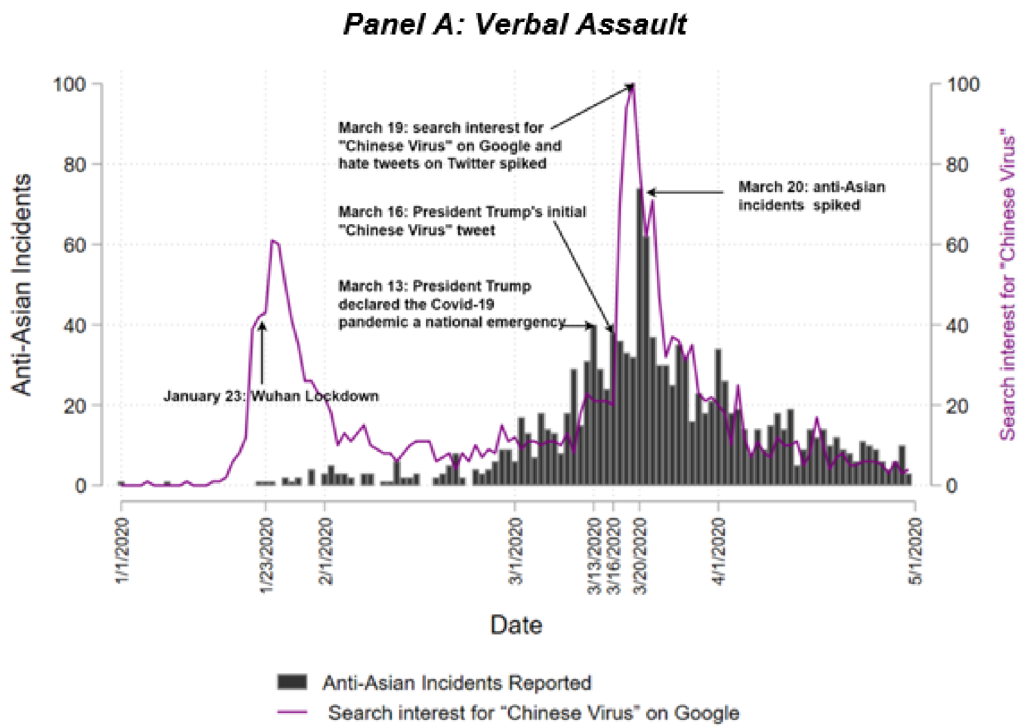


Figure B.5: Daily Anti-Asian Incidents, Verbal Assaults and Shunning



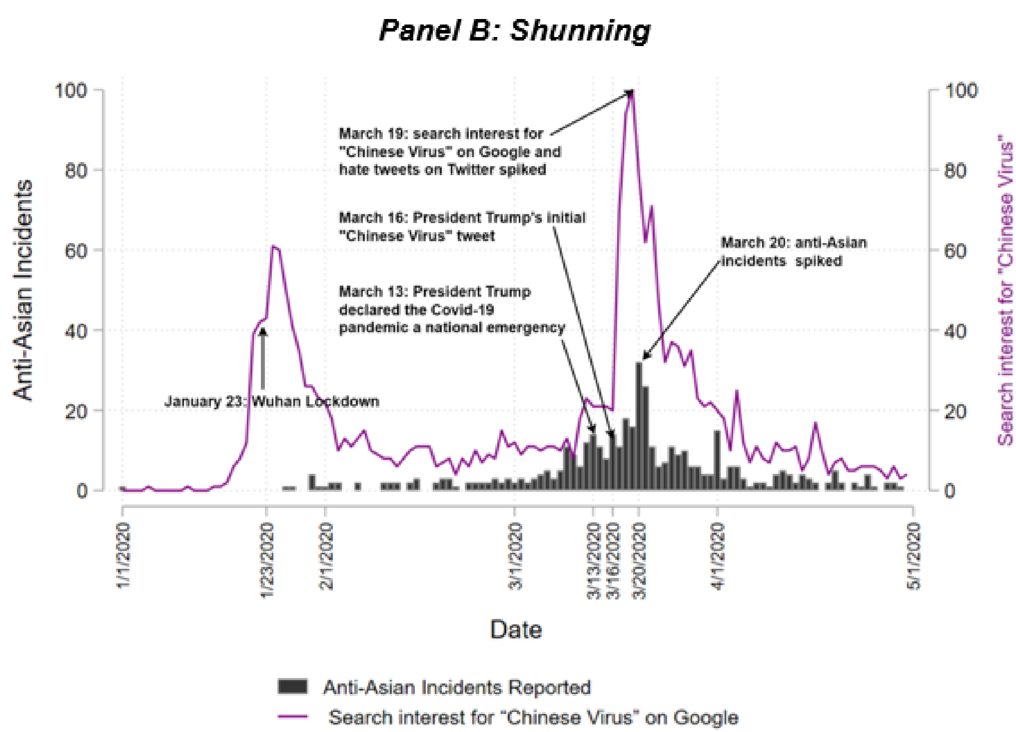
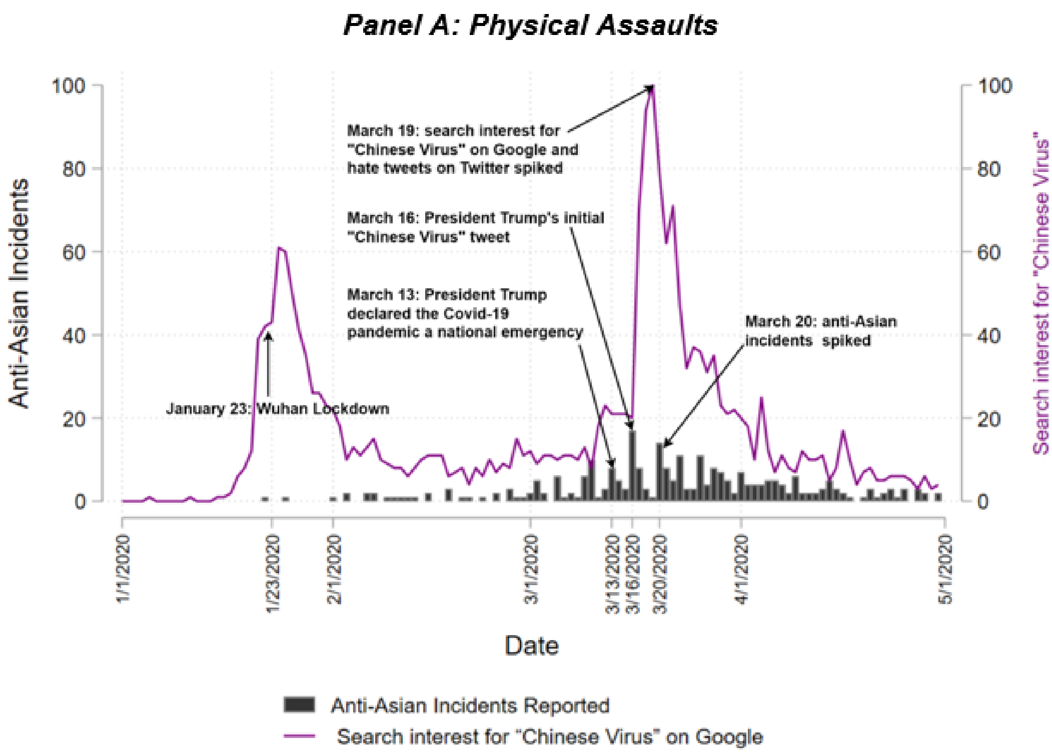


Figure B.6: Daily Anti-Asian Incidents, Physical Assaults and Workplace Discrimination



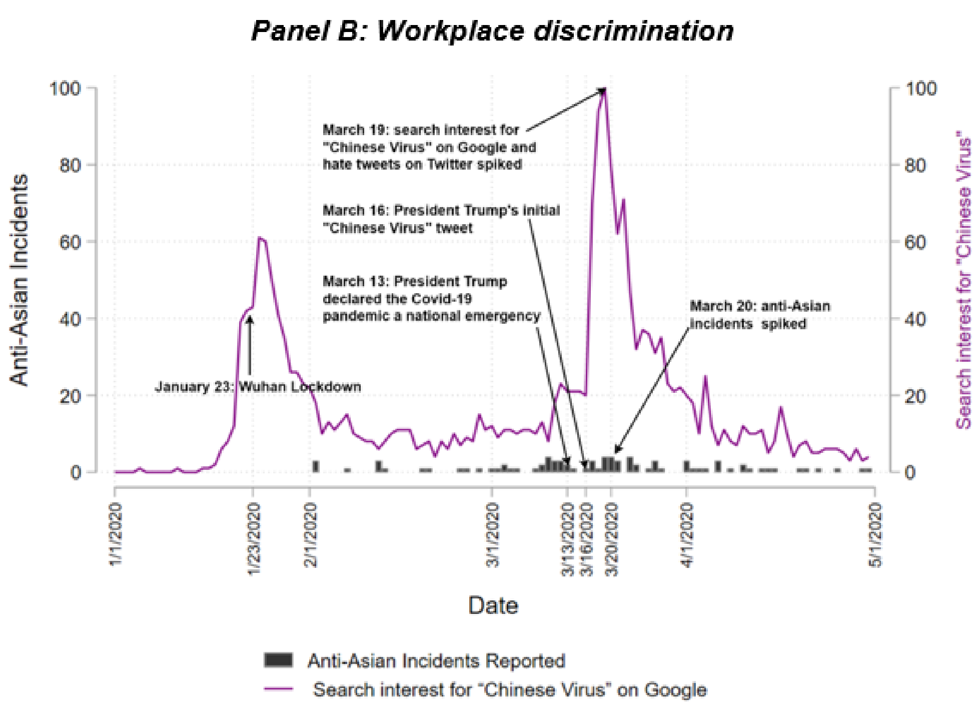
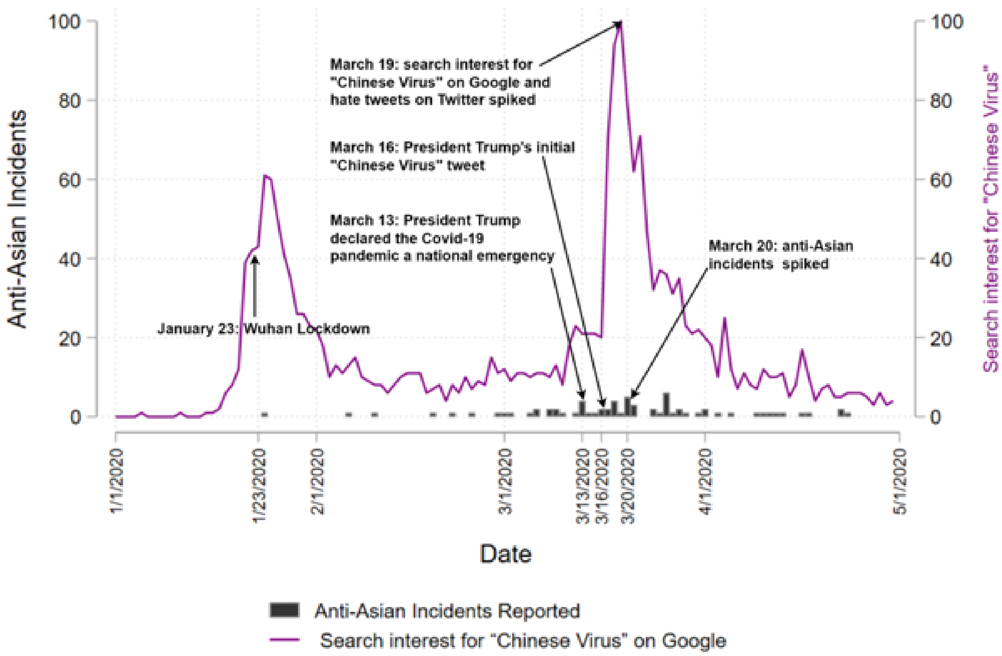


Figure B.7: Daily Anti-Asian Incidents, Physical Assaults and Workplace Discrimination

Panel A: Refusal of Service



Panel B: Online misconduct

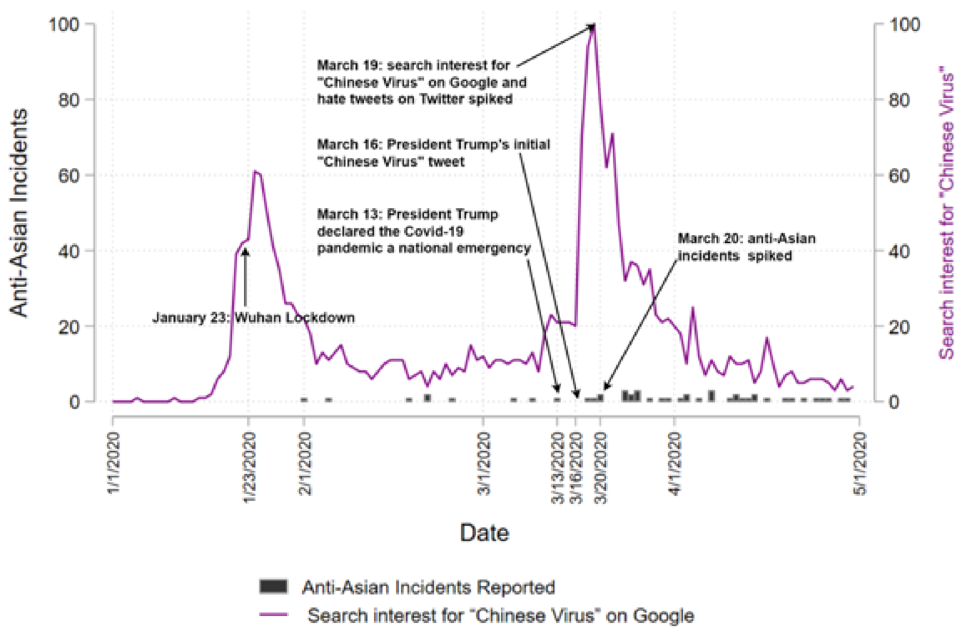


Figure B.8: Daily Anti-Asian Incidents, Other Incident Types

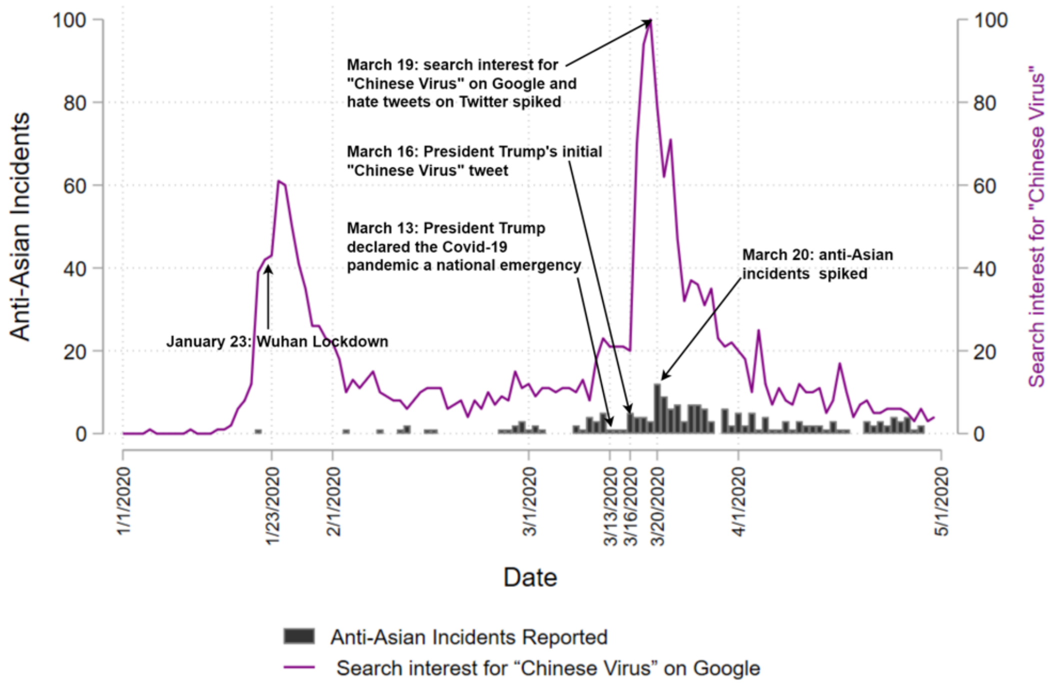


Figure B.9: Daily Anti-Asian Incidents, Other Incident Types

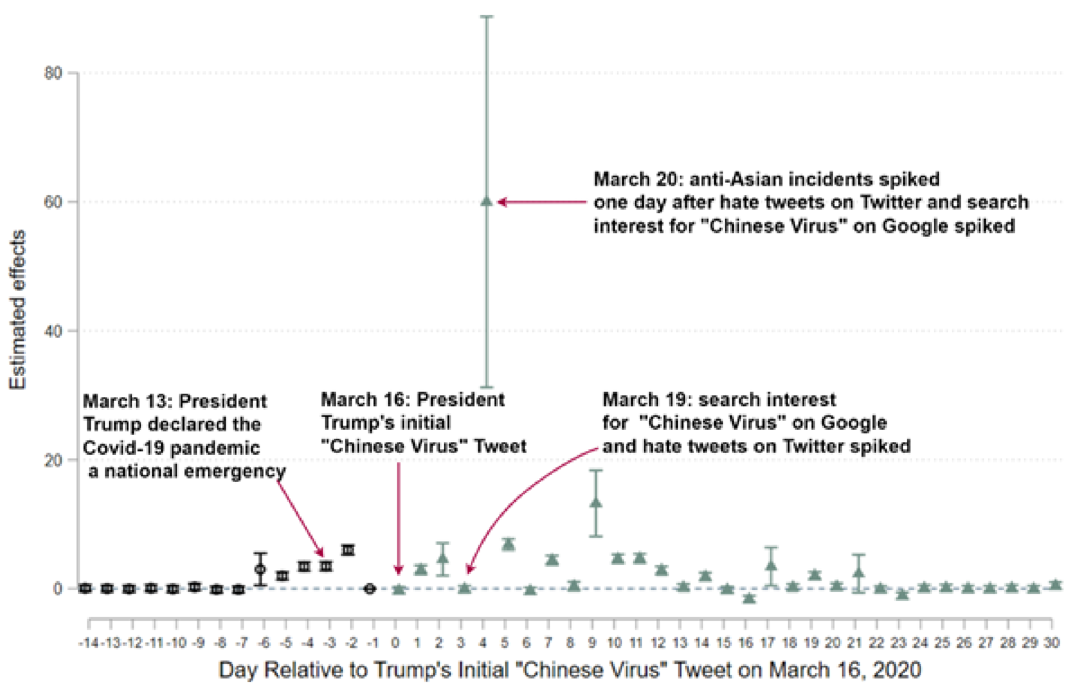
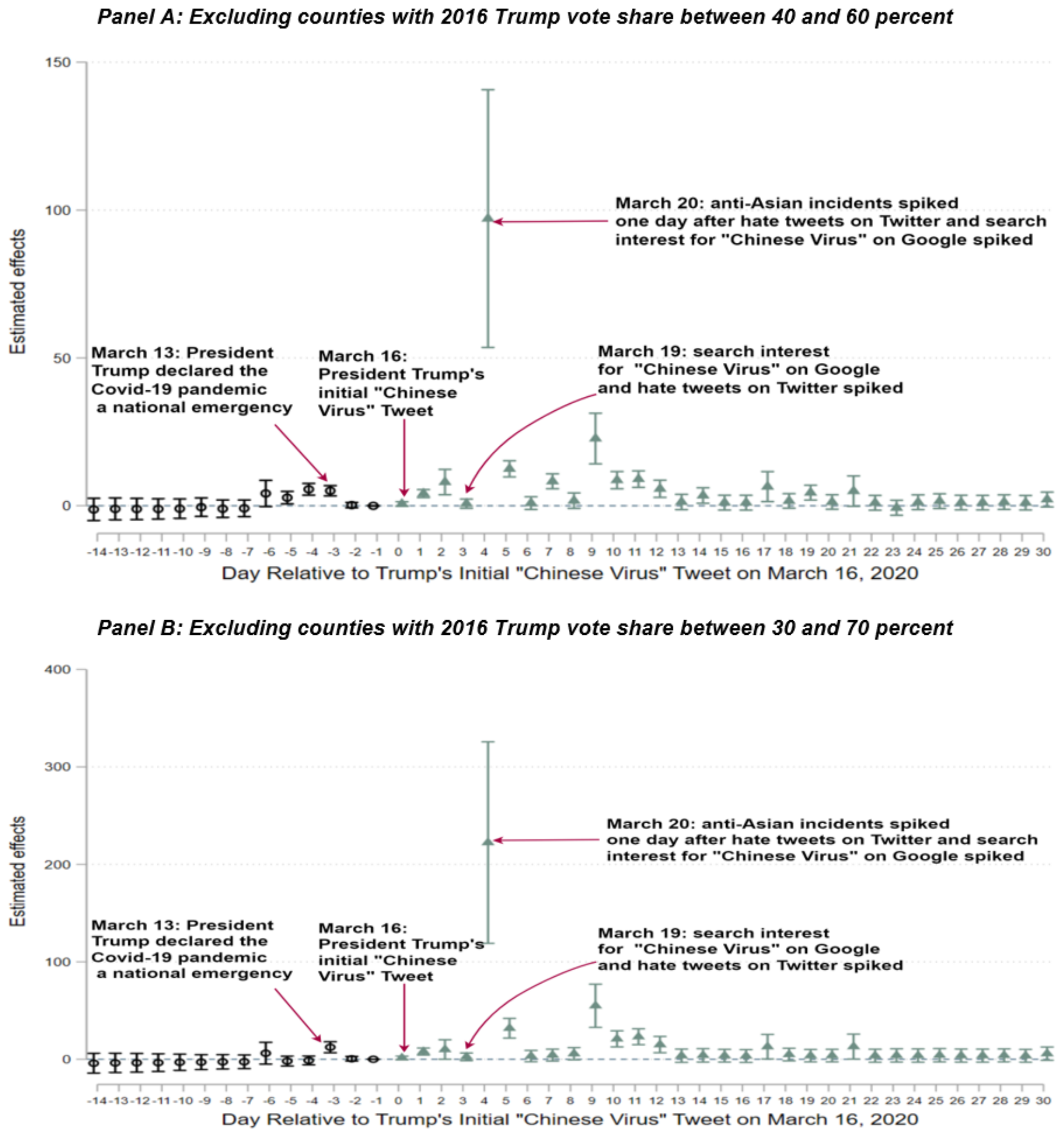
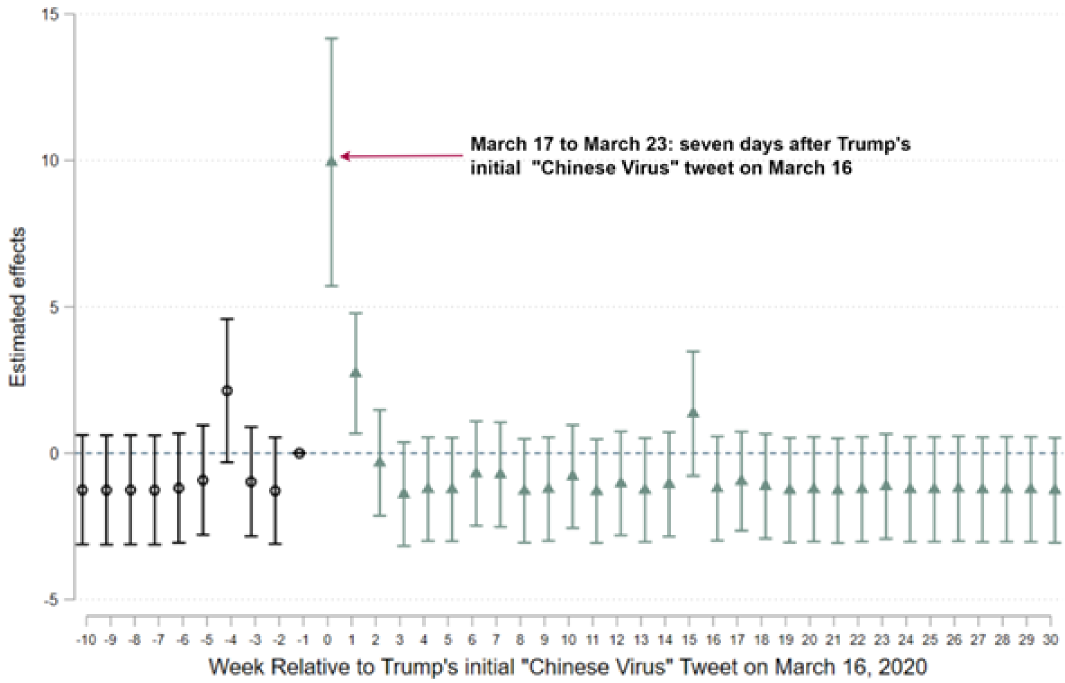


Figure B.10: Estimated Effects on Anti-Asian Incidents in Trump- vs Clinton-supported Counties, omitting Counties with Similar Levels of Support



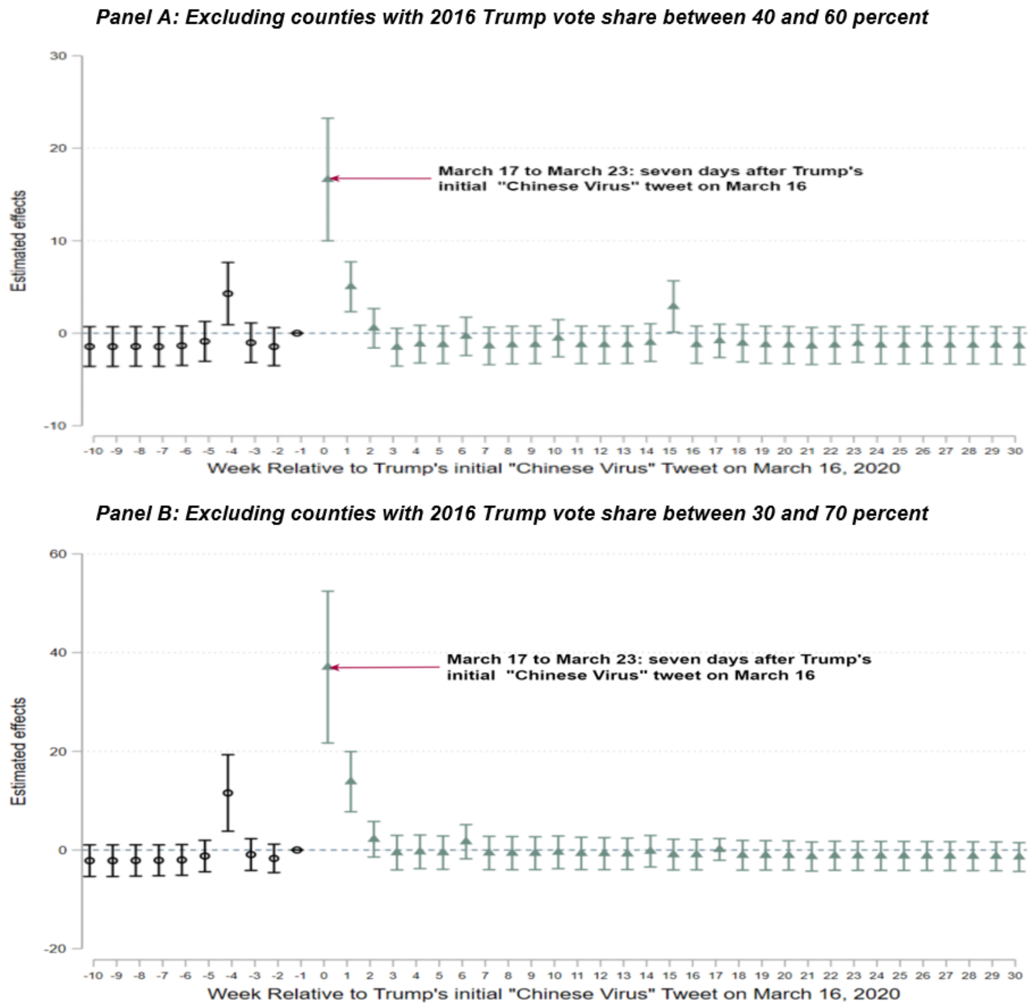
Notes: These figures plot the estimated effects of Trump's initial "Chinese virus" tweet on anti-Asian incidents in counties that supported Trump in 2016 versus those that supported Clinton. Panel A excludes counties for which Trump received between 40 and 60 percent of the 2016 presidential vote. Panel B excludes counties for which Trump received between 30 and 70 percent of the 2016 presidential vote. Estimates control for the logarithm of the total number of Covid-19 cases plus one, county fixed effects, and date fixed effects. The outcome variable is the number of reported anti-Asian incidents per 100,000 Asian residents. Data, restricted to incidents 3/2/20-4/15/20, are from the Stop AAPI Hate database. Confidence intervals are based on two-way standard-error estimates allowing for clustering within counties over time and across counties on the same date A. C. Cameron et al. (2011).

Figure B.11: Estimated Effects on Anti-Asian Incidents in Trump- vs Clinton-Supported Counties, Using Data Aggregated into 7-day Periods



Notes: These figures plot the estimated effects of Trump's initial "Chinese virus" tweet on anti-Asian incidents in counties that supported Trump in 2016 versus those that supported Clinton. Estimates control for the logarithm of the total number of Covid-19 cases plus one, county fixed effects, and week fixed effects. The outcome variable is the number of reported anti-Asian incidents per 100,000 Asian residents. Data, restricted to incidents 1/1/20-10/19/20, are from the Stop AAPI Hate database and are aggregated to 7-day periods (i.e., "weeks") relative to Trump's initial "Chinese Virus" tweet on March 16, 2020. For example, week "0" includes March 17 through March 23, 2020, the 7-day period starting the day after Trump's initial tweet. Confidence intervals are based on two-way standard-error estimates allowing for clustering within counties over time and across counties on the same week A. C. Cameron et al. (2011).

Figure B.12: Estimated Effects on Anti-Asian Incidents in Trump- vs Clinton-supported Counties, omitting Counties with Similar Levels of Support , Using Data Aggregated into 7-day Periods



Notes: These figures plot the estimated effects of Trump's initial "Chinese virus" tweet on anti-Asian incidents in counties that supported Trump in 2016 versus those that supported Clinton. Panel A excludes counties for which Trump received between 40 and 60 percent of the 2016 presidential vote. Panel B excludes counties for which Trump received between 30 and 70 percent of the 2016 presidential vote. Estimates control for the logarithm of the total number of Covid-19 cases plus one, county fixed effects, and week fixed effects. The outcome variable is the number of reported anti-Asian incidents per 100,000 Asian residents. Data, restricted to incidents 1/1/20-10/19/20, are from the Stop AAPI Hate database and are aggregated to 7-day periods (i.e., "weeks") relative to Trump's initial "Chinese Virus" tweet on March 16, 2020. For example, week "0" includes March 17 through March 23, 2020, the 7-day period starting the day after Trump's initial tweet. Confidence intervals are based on two-way standard-error estimates allowing for clustering within counties over time and across counties on the same date A. C. Cameron et al. (2011).

Figure B.13: Estimated Effects on Anti-Asian Incidents in Trump- vs Clinton-supported Counties, Verbal Assault and Shunning

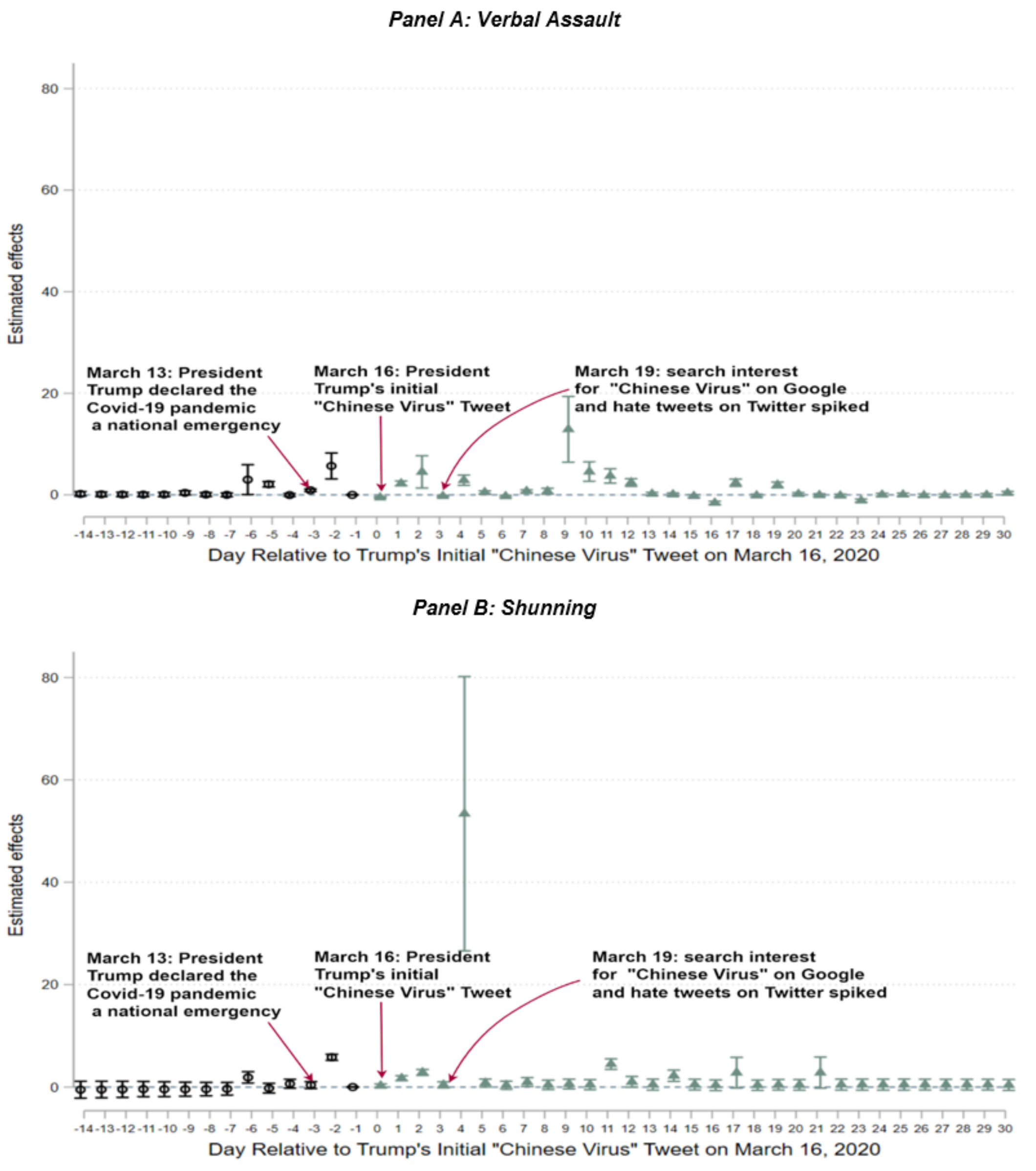


Figure B.14: Estimated Effects on Anti-Asian Incidents in Trump- vs Clinton-supported Counties, Refusal of Service and Online Misconduct

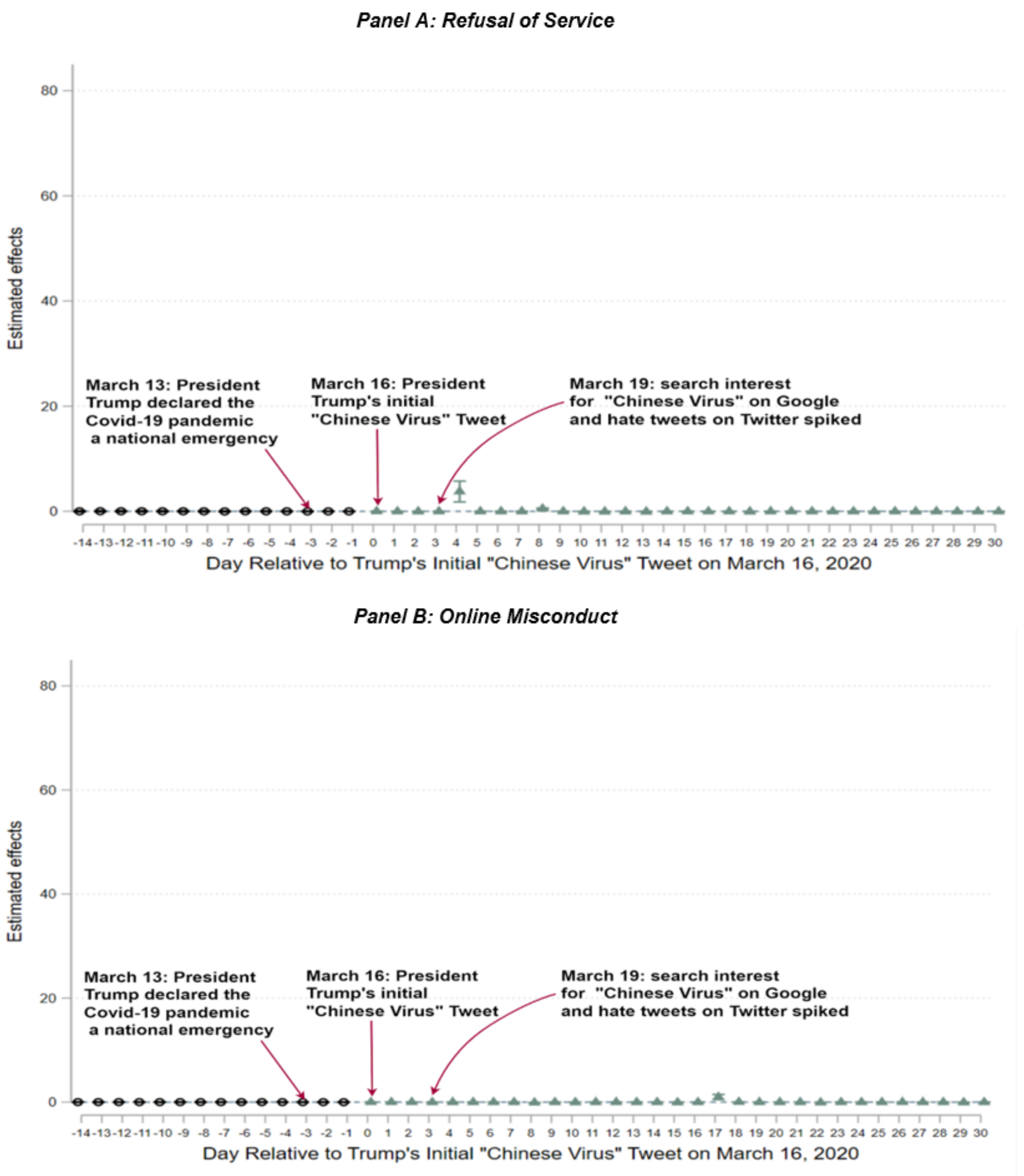


Figure B.15: Estimated Effects on Anti-Asian Incidents in Trump- vs Clinton-supported Counties, Physical Assault, Workplace Discrimination, and Other Incidents

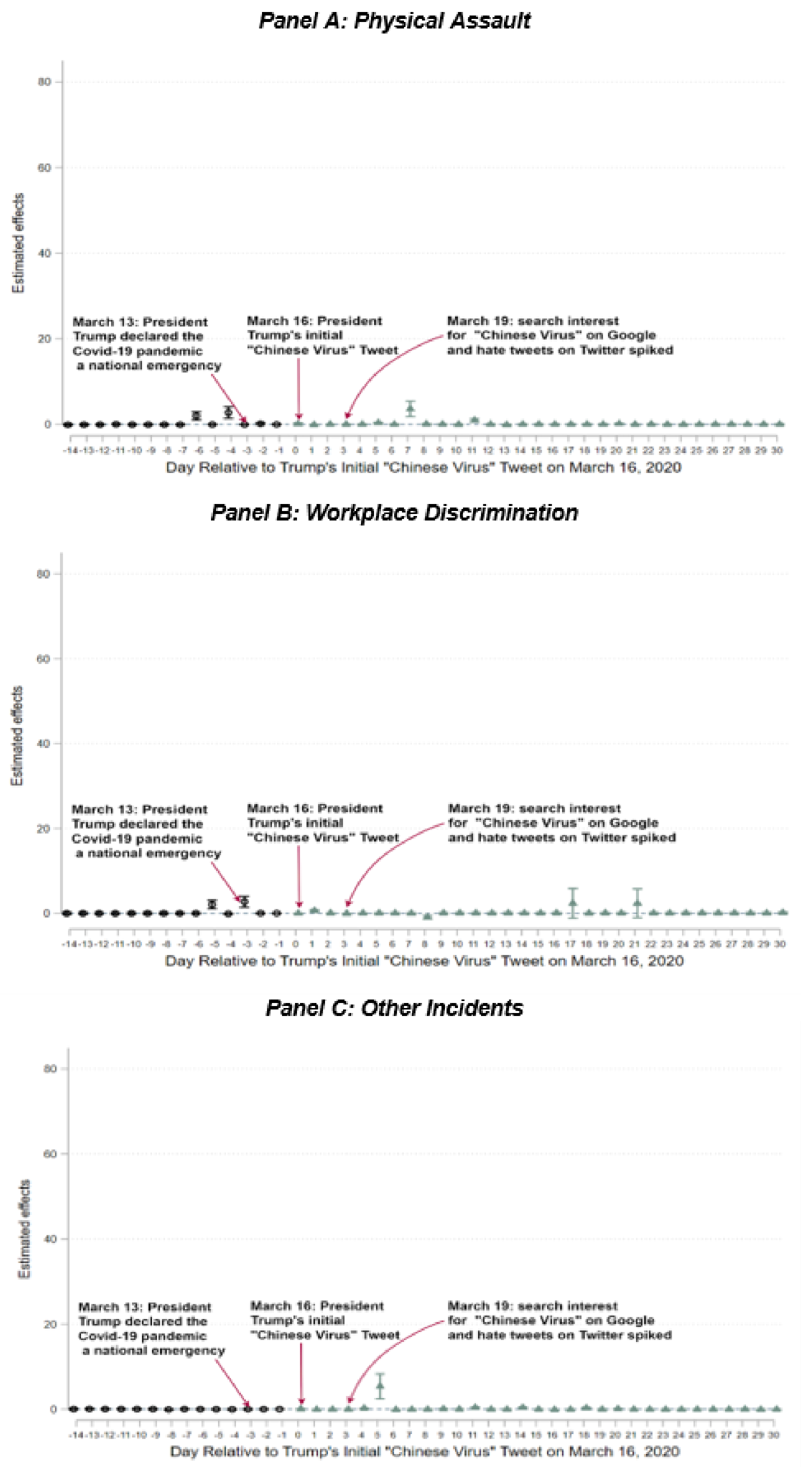
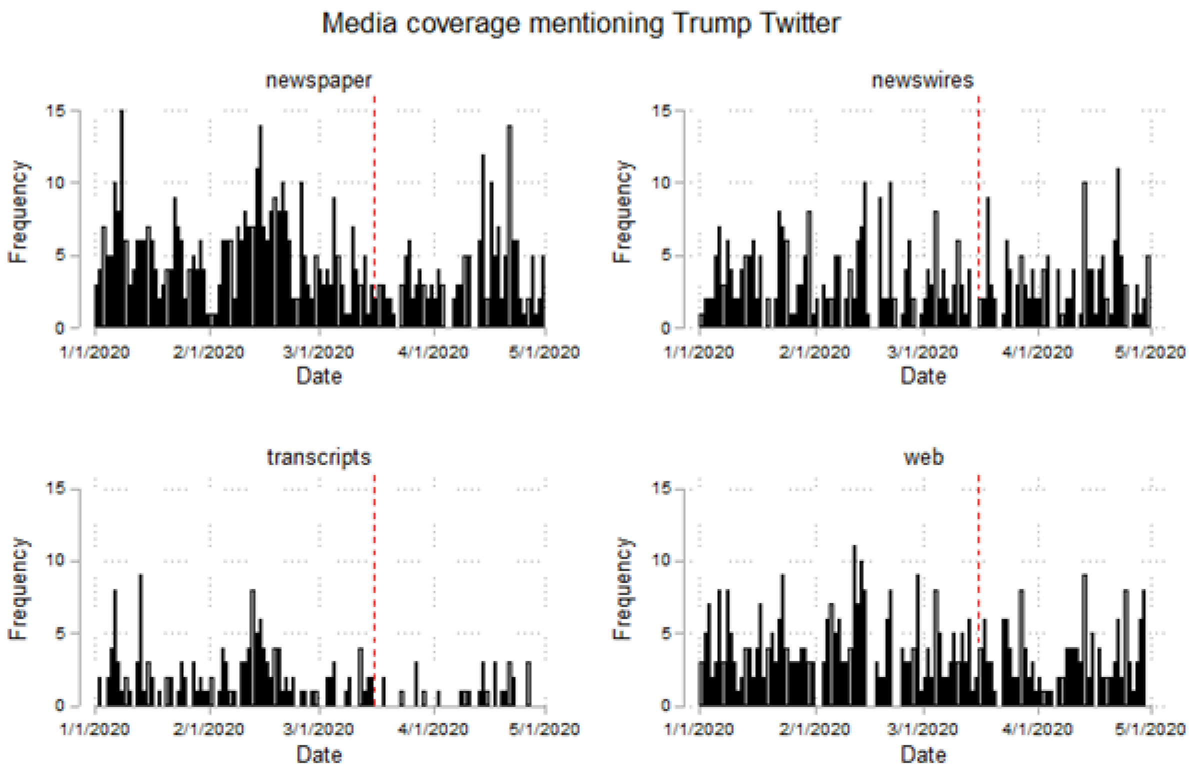
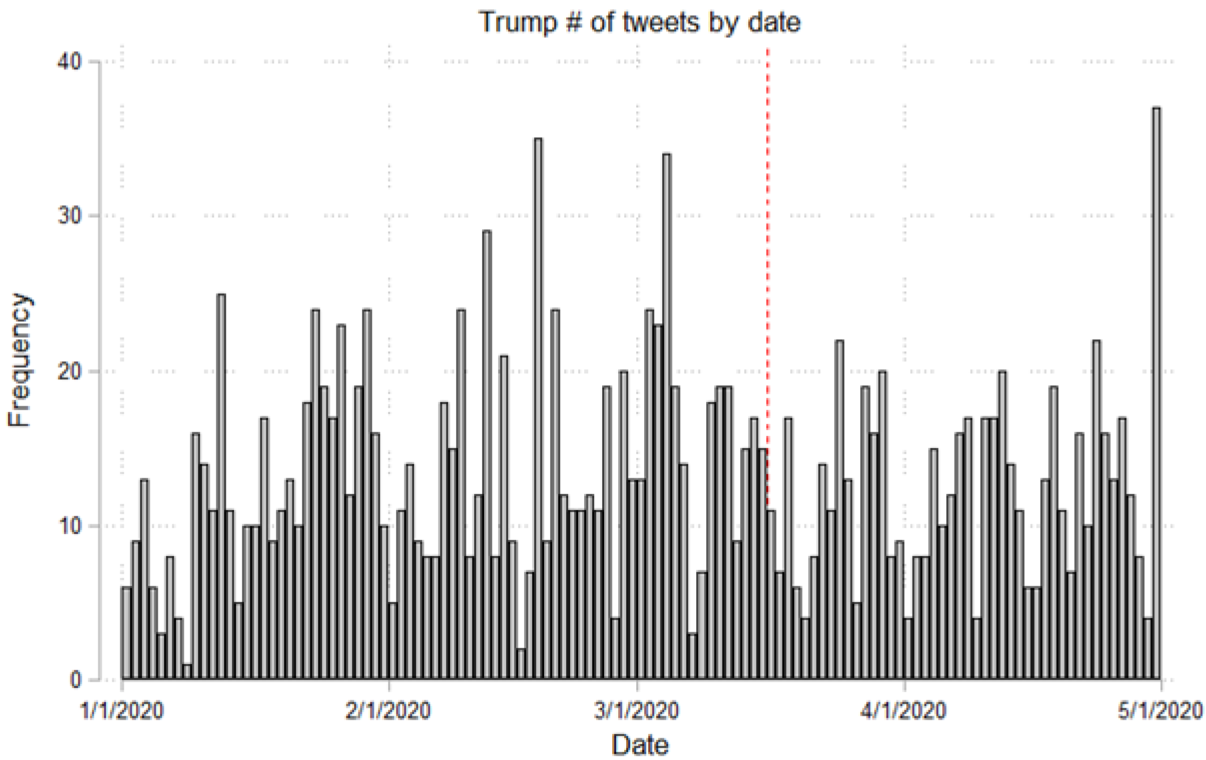


Figure B.16: Media Outlet Mentions of "Trump" and "Twitter"



Notes: The red vertical lines are drawn the date of Trump's first "Chinese virus" tweet (3/16/2020).

Figure B.17: Trump's Twitter Activity by Date



Notes: The red vertical line is drawn the date of Trump's first "Chinese virus" tweet (3/16/2020).

Table B.1: Incidents and Anti-Asian COVID-19 Tweets (per 100,000 Asian Residents), Average Daily Means Based on County-day Observations

	14 days before Trump's Initial "Chinese Virus" Tweet		14 days after (inclusive) Trump's Initial "Chinese Virus" Tweet		Day of spike	
	Clinton	Trump	Clinton	Trump	Clinton	Trump
Incidents	0.152	1.392	0.444	7.456	0.449	60.370
Anti-Asian Covid-19 Tweets	1.790	2.971	3.634	7.823	9.289	22.389

Notes: Anti-Asian COVID-19 tweets spiked to their highest point three days after Trump's first use of "Chinese Virus" and anti-Asian incidents spiked to their highest point the following day.

APPENDIX C

CHAPTER III APPENDIX

C.1 Screenshots

Figure C.1: Real-effort Task: Encoding

Citizen task: Sequence 1 Round 1

Time left to complete this page: **0:07**

- You are given a table that converts each letter in the alphabet into a unique number.
- You are given a random string of letters. Your job is to translate the string into numbers based on the table you are given.
- For example, cjq would be translated to 31017

a	b	c	d	e	f	g
1	2	3	4	5	6	7
h	i	j	k	l	m	
8	9	10	11	12	13	
n	o	p	q	r	s	t
14	15	16	17	18	19	20
u	v	w	x	y	z	
21	22	23	24	25	26	

Your job is to encode the following string of letters:

g

Please enter your answer in the box below:

Your job is to encode the following string of letters:

k

Please enter your answer in the box below:

Your job is to encode the following string of letters:

l

Please enter your answer in the box below:

Figure C.2: Real-effort Task: Counting Zeros

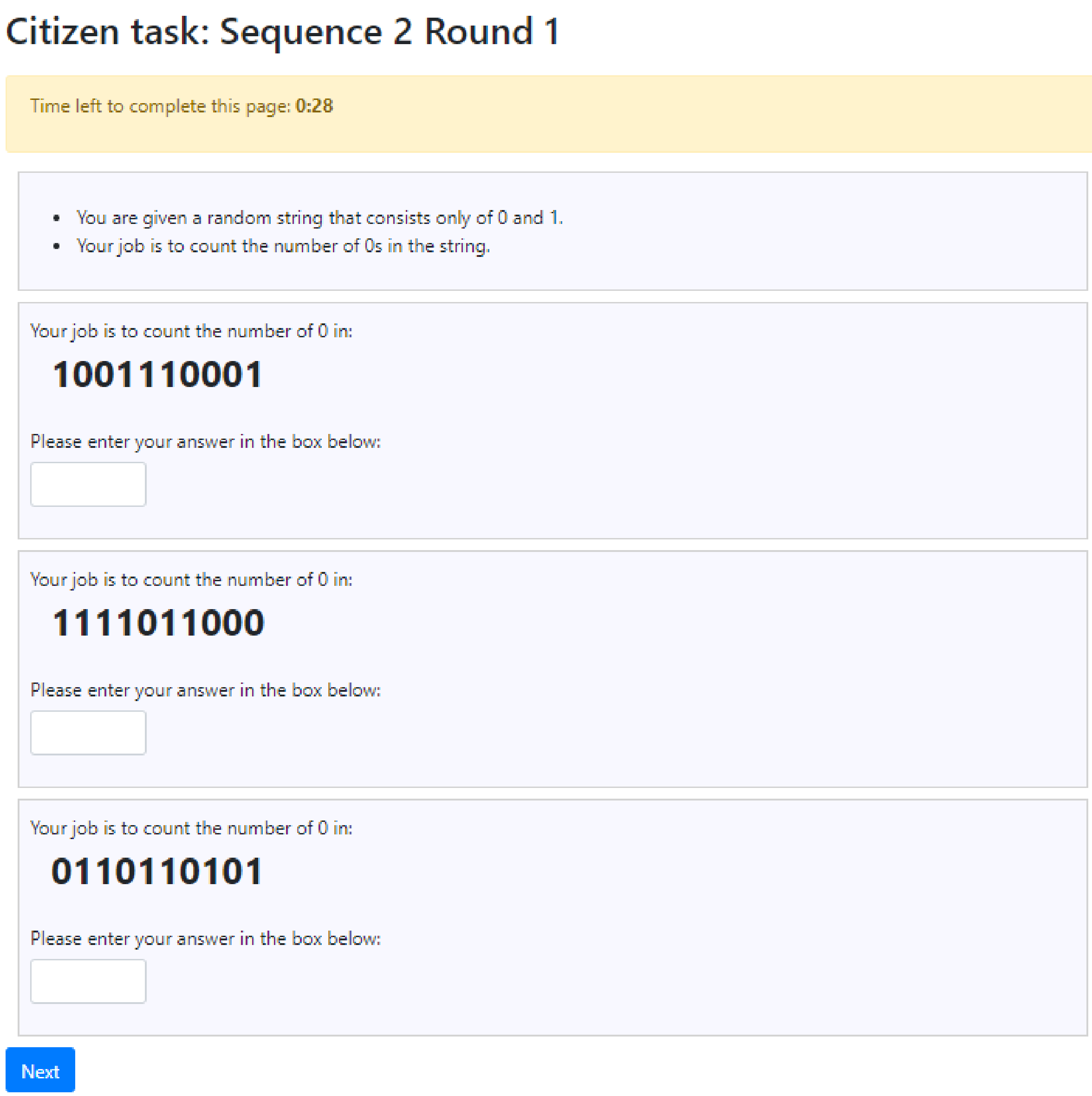
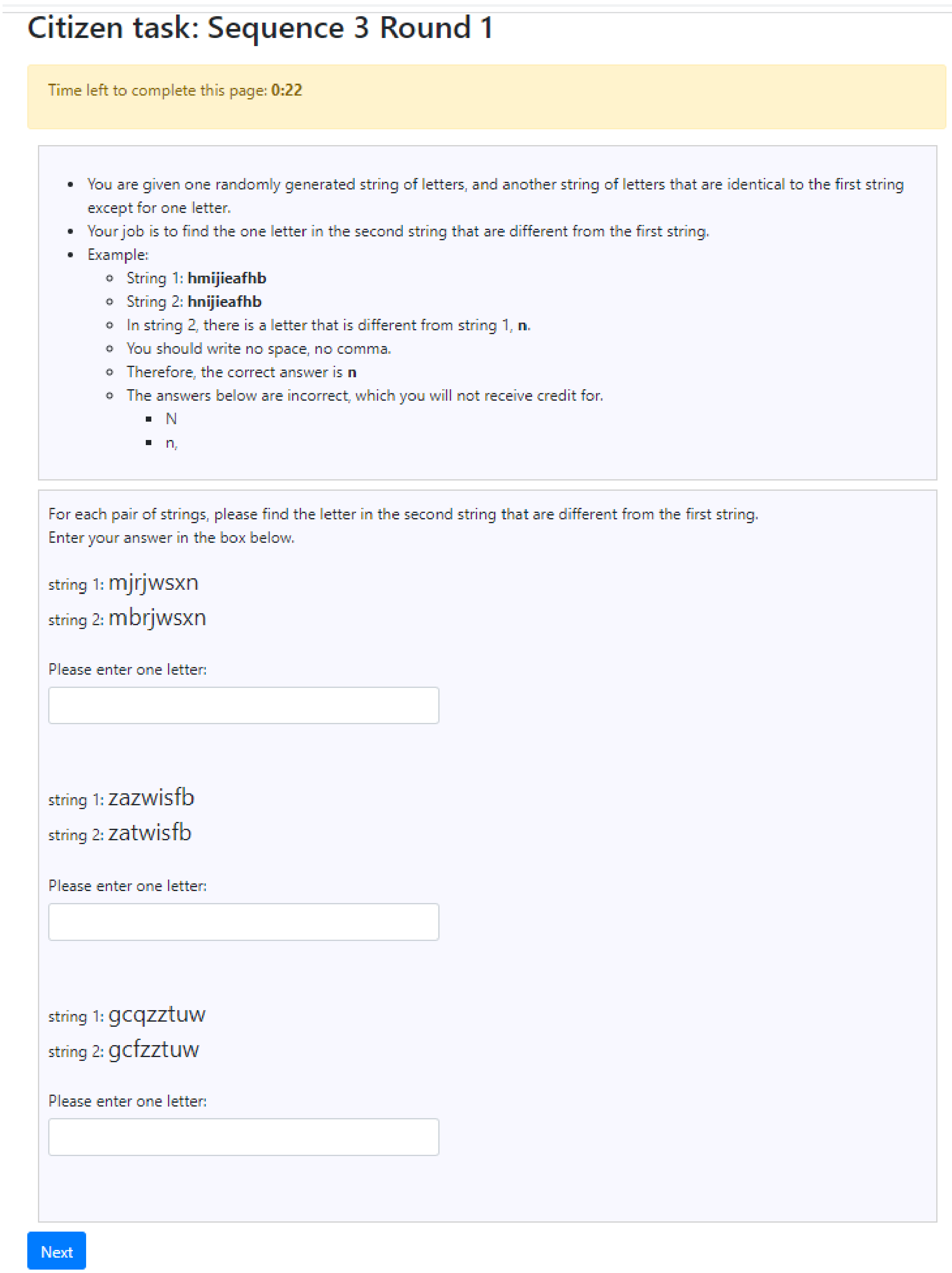


Figure C.3: Real-effort Task: Finding Letters



Citizen task: Sequence 3 Round 1

Time left to complete this page: **0:22**

- You are given one randomly generated string of letters, and another string of letters that are identical to the first string except for one letter.
- Your job is to find the one letter in the second string that are different from the first string.
- Example:
  - String 1: **hmijieafhb**
  - String 2: **hnijieafhb**
  - In string 2, there is a letter that is different from string 1, **n**.
  - You should write no space, no comma.
  - Therefore, the correct answer is **n**
  - The answers below are incorrect, which you will not receive credit for.
    - N
    - n,

For each pair of strings, please find the letter in the second string that are different from the first string.
Enter your answer in the box below.

string 1: mjrjwsxn
string 2: mbrjwsxn

Please enter one letter:

string 1: zazwisfb
string 2: zatwisfb

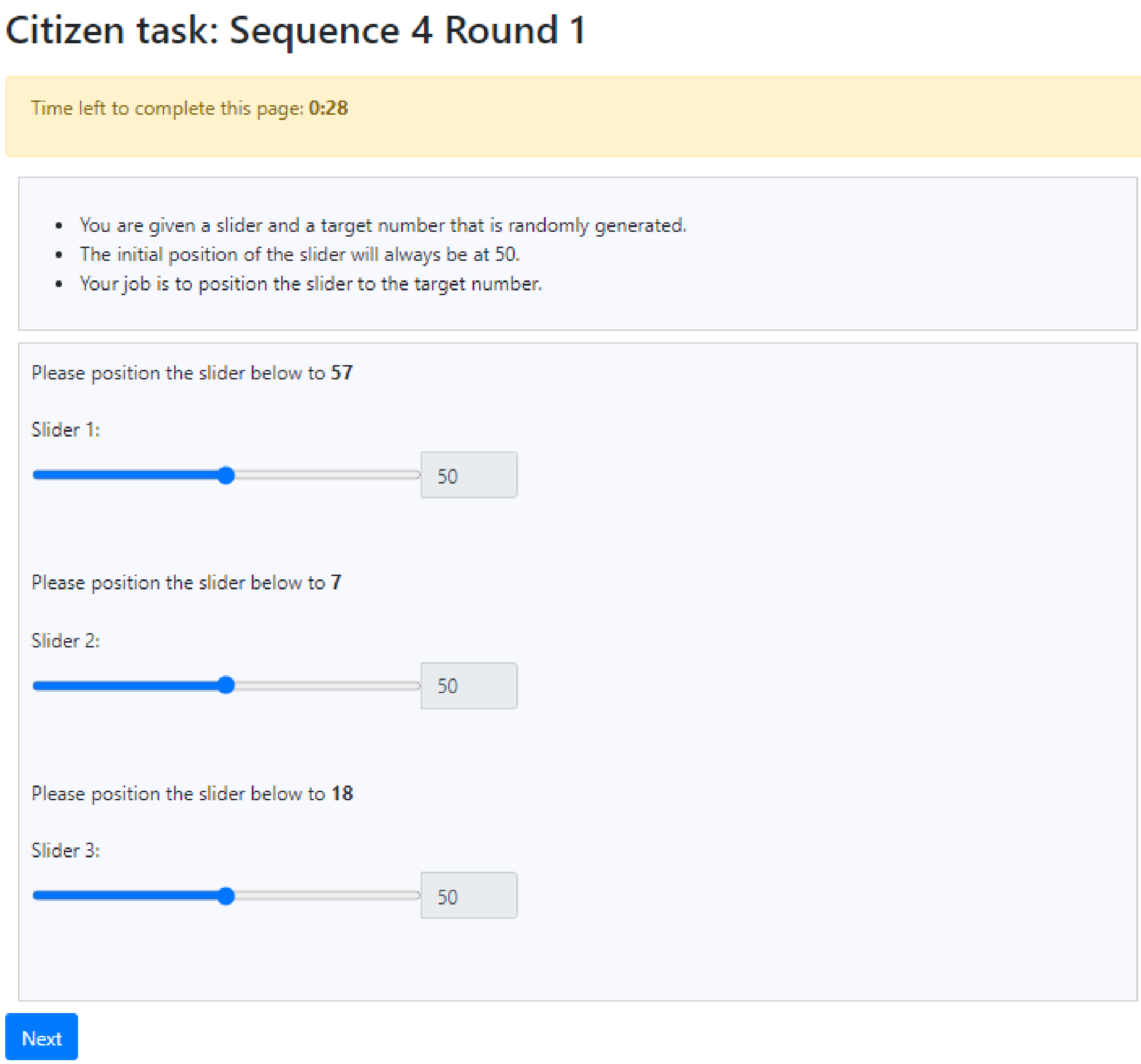
Please enter one letter:

string 1: gcqzztuw
string 2: gcfzztuw

Please enter one letter:

Next

Figure C.4: Real-effort Task: Sliders



C.2 Chat Example

Typical reasoning in the **Honest** Committee Treatment

- yeah i feel like if we do it everytime itll be lame later if we end up being citizens
- hey i forgot to mention the two people that stay in the committee make sure that the new member knows that more money can be earned overall if you choose to distruibute to the public everytime
- newbie its better to redistrubte bc the potential for income in the future is greater for each of us
- and if they decide to not work on it, we get screwed
- everyone should get the money
- It's just better for everyone if we distribute it evenly
- yeah but it would be better for society if we split it with the whole group

Typical reasoning in the **Corrupt** Committee Treatment

- yall keep stealin!
- keep money let citizen be poor
- guys if yall redistribute the citizens get more than us
- because citizens make more
- because we would make up for what we are losing compared to the citizens
- powerrrrrr
- been here three rounds the power got to my head now
- let's keep the money earned to ourselves
- They don't know that we're keeping it. For all they know, we're bad at the questions or the fund didn't triple due to probability

C.3 Instructions

"A study of individual and group decision-making"

INSTRUCTIONS (T1)

(READ ALOUD)

General instructions

Thank you all for coming today. You are here to participate in an experiment. In addition to a $10 participation fee, you will be paid any money you accumulate from the experiment. You will be paid privately, by Venmo or PayPal, at the conclusion of the experiment.

The number that you have found on your desk is your *identification number* in the experiment. We will not ask you to write down your name at any time during this experimental session. No one, including the experimenter, will link your name to the decisions you made in the experiment.

The experiment will consist of two parts. Part 1 is made of four different activities. Part 2 is made of one activity that will be repeated for a number of rounds. The instructions for each activity will be provided on your screen. Earnings during the experiment will be denominated in Experimental Currency Units, or ECU. At the end of the session, one activity from Part 1 and one round from Part 2 will be randomly selected for payment and your earnings in that stage will be converted to dollars at the exchange rate of **$1 for 12 ECU**. After participating in all the stages of the experiment you will be asked to complete a brief questionnaire. You will then be paid the money your earned in the experiment.

This study has been reviewed and approved by the Texas A&M Human Subjects Committee. *If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment.* Participants intentionally violating these rules may be asked to leave the experiment and may not be paid.

Please read the information sheet that you have been given. Please raise your hand if you have any question about any of the information on the information sheet. We will proceed with the experiment once everyone is ready. Please note that you have found on your desk a handout regarding Part 2 of the experiment. We will go through the handout together at the conclusion of Part 1. Please do not read the handout until then.

[Start program]

On your screen it should say please enter your participant label. I am going to ask you to enter a 4-digit number, it is very important that you enter the number correctly. The first two digits are 01, the last two digits are your station number, which is the number that you have been given, and is also on your cubicle – on your screen. This will allow us to pay you your earnings at the end of the session.

Sometimes on your screen it will say waiting for the other participants, it means other people are still making decisions, so please wait patiently.

Handout - PART 2
(READ ALOUD)

We are about to start Part 2 of the Experiment. You can take out the handout. You will also find the same information on your screen. I am going to read it out loud. Please follow along. You will be playing the activities of Part 2 multiple rounds. One of these rounds will be randomly selected for payment at the conclusion of the experiment.

In this part, you will be a member of a society made of 8 participants, you and 7 others. As before, you will not know who the other participants in your society are. You will stay in the same society until the end of the experiment.

At the very beginning, 3 participants will be assigned the role of "Committee Members" and the remaining 5 participants will be assigned the role of "Citizens." Committee members and citizens will engage in different tasks during the experiment.

- The tasks of the **5 Citizens**:
  - In each round, **Citizens** will receive a fixed wage of 100 ECU and will engage in an activity for 30 seconds. The activity will consist of solving some simple tasks.
  - If a citizen successfully completes 3 tasks in 30 seconds, he or she will generate 50 ECU in additional earnings.
  - However, the Citizen can only keep 36% of the additional earnings, which is 18 ECU. The remaining 64% of citizens' additional earnings, i.e., 32 ECU per citizen, will be deposited into a public fund. Since there are 5 citizens in a society, this means citizens can generate up to 160 ECU to be deposited in the public fund (5×32=160).
  - Depending on the actions of the 3 Committee members and on luck, the public fund could be tripled, in which case the Citizens could receive back an equal share of the fund, i.e., an amount equal to 1/8 of the total tripled money in the fund.

- The tasks of the **3 Committee Members:**
  - In each round, they receive a wage of 80 ECU, and are in charge of the public fund;
  - They engage in a task and:
    - If they cannot successfully complete the task, the money in the public fund is lost;
    - If they are jointly successful in the task, the public fund is tripled with probability 80% or lost with probability 20%;
  - If the public fund is successfully tripled, the 3 committee members have the task of redistributing the money equally among themselves and the citizens.
  - However, they can instead jointly decide to keep the money and divide only among the 3 of them. The outcome will be decided by the 3 committee members via majority voting.
  - Importantly, only the committee members will know if they were successful in the task, and whether the public fund got tripled or if the money is lost.

- This means that, if Citizen do not receive dividends from the public fund, they will not know if it was because the public fund was not successfully tripled, or because the Committee members decided to keep the money in the fund for themselves.
- The committee members will be able to talk with each other via chat for 2 minutes at the beginning of round 1 and again at the beginning of round 6.

Citizens and Committee members will make the same decision for 10 rounds. At the end of each round, Citizens will be informed about the amount they received from the public fund, if any.

At the end of the 10 rounds, <u>a new sequence of 10 rounds</u> will begin. However, one member of the Committee will be <u>randomly chosen to step down</u> and <u>a Citizen will replace him or her</u>.

Each Citizen will be asked if they want to be part of the Committee, and the replacement will be randomly chosen among those who have expressed an interest in serving in the Committee. The new Committee members will make decisions for the next 10 rounds. At the end of the second sequence of 10 rounds, another Committee member will be chosen to step down, and a new Citizen will be chosen to replace him or her, for the next 10 rounds. Finally, at the end of the third sequence of 10 rounds another Committee member will step down and will be replaced by a new Citizen for the final 10 rounds. Part 2 of the experiment will conclude after the 4 sequences of 10 rounds have been completed.

To summarize, here are the potential earnings of Citizens and Committee members in each round of Part 2:

- **Citizens** get a wage of **100 ECU**, and, on top of that, can earn 18 ECU if they successfully complete 3 simple tasks. Moreover, they can earn 1/8 of the money in the public fund, if the fund is successfully tripled. If the fund is not tripled, Citizens earn 118 ECU;

- **Committee members** get a wage of **80 ECU** and, on top of that, can earn either 1/8 or 1/3 of the money in the public fund – depending on their joint decisions – if the fund is successfully tripled. If the fund is not tripled, Committee Members earn 80 ECU.

Are the roles of Citizen and Committee Member clear?

In the next screen, you will see some comprehension questions that are designed to test your understanding, after you answer the questions correctly, you will know if you will assume the role of Citizen or Committee member in the first Sequence of 10 rounds. You will then be shown the instructions for Part 2. Round 1 will then begin.

If at any time you have any questions about the Part 2 of the experiment, please raise your hand.